The Cognitive and Neural Mechanisms of Joint Attention: A Second Person Approach

Nathan Caruana, B Psychology (Honours 1st Class)

Department of Cognitive Science

ARC Centre of Excellence in Cognition and its Disorders

Perception in Action Research Centre

Faculty of Human Sciences

Macquarie University, Sydney, Australia

This dissertation is presented for the degree of Doctor of Philosophy (PhD)

August, 2015

Table of Contents

Summary

This project investigated the cognitive and neural mechanisms of joint attention in typical development and in high-functioning autism. A novel gaze-contingent virtual reality tasks was developed and implemented in a number of studies using a range of techniques, including functional magnetic resonance imaging (fMRI), event-related potentials (ERP) and eye tracking.

The first aim of this project was to develop an experimental joint attention paradigm that (1) captured both the initiating and responding functions of joint attention, (2) could be applied in both behavioural and neurophysiological experiments, (3) provided full experimental control over non-social task demands, and (4) established an ecologically valid context for joint attention interactions.

Chapter 2 of this dissertation reviewed the current approaches for measuring joint attention in experimental settings and introduced a new virtual reality paradigm of joint attention that achieves both experimental control and ecological validity. Critical issues associated with the measurement of joint attention are discussed.

The second aim of this project was to use this paradigm to investigate the neural and cognitive mechanisms that support joint attention in typical development and in autism.

In **Chapter 3**, fMRI was used to investigate the neural correlates that were unique and common to initiating and responding to joint attention bids in 13 adults with typical development. A right-lateralised frontotemporoparietal network was found to be common to both initiating and responding to joint attention bids and comprised 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. Compared to responding to joint attention bids, initiating joint attention was associated with additional activation of the MFG, IFG, TPJ and precuneus.

In **Chapter 4**, eye-tracking was used to investigate joint attention performance in 17 adults with high-functioning autism and 17 adults with typical development (controls). Individuals with autism made significantly more errors than controls when initiating and responding to joint attention bids.

In **Chapter 5**, another virtual reality paradigm was developed and employed in an event-related potential study which investigated the time course of neural processes associated with evaluating self-initiated joint attention bids. In a sample of 19 participants with typical development, centro-parietal P350 and P500 ERPs were significantly larger when gaze shifts resulted in the avoidance, rather than the achievement of joint attention. This P350 and P500 morphology was absent in a second sample of 19 participants who completed a non-social version of the same task in which arrows replaced the gaze of the virtual character.

In **Chapter 6**, the same paradigm was used to investigate whether the P350 effect observed in **Chapter 5** was influenced by participants' beliefs of their virtual partner's agency. The P350 effect was only observed in participants who believed that their partner was controlled by a human (n = 19), and not in a second group of individuals who were informed that the virtual character was controlled by a computer program (n = 19).

Declaration

I, Nathan Caruana, certify that the work in this dissertation entitled "The Cognitive and Neural Mechanisms of Joint Attention: A Second Person Approach" has not been previously submitted for a degree, nor has it been submitted as part of the requirements for a degree to any university or institution other than Macquarie University.

I also certify that the dissertation is an original piece of research and it has been written by me. Any help and assistance that I have received in my work, and the preparation of the dissertation itself has been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the dissertation.

The research presented in this dissertation was approved by the Macquarie University Human Ethics Review Committee, reference number: **5201200021**. This was also ratified by the University of Western Australia Ethics Committee, reference number: **RA/4/1/7185**.

Signed:

Nathan Caruana (Student ID: 41479777)

August, 2015

I would also like to thank Lesley McKnight, Robin Blumfield, Marcus Ockenden, Craig Richardson, Teri Roberts and Lisa Yen for their friendship first, and for the many times I sought their counsel to navigate the personal, administrative and technical challenges that I encountered during my candidature.

Finally, I must acknowledge my most valuable resource – my family. Words cannot express the gratitude and love that I have for them. To my brother Rodney, his wife Josephine, and my sisters Julie and Tania – I thank them for their endless support, in everything. They have always reminded me of what is important in life and helped me to refocus my priorities at times when I became consumed with work or study.

To my parents, Paul and Marthese – they are the smartest, most generous, and loving people I know. I owe every single achievement I own to them. I thank them for selflessly supporting my ambitious. They have sheltered me, fed me, and encouraged me my entire life, and I am eternally grateful to them both.

To my wife Catherine – I would like to thank her for believing in me. From the beginning to the end of my PhD she has patiently listened, gently counselled, and fiercely encouraged me. Cath has ensured that every day begins and ends with happiness, and for that I will always be in her debt.

Chapter 1

General Introduction

General Introduction

Introduction

Humans are innately social creatures with a biological need for social interaction (Baumeister & Leary, 1995). We seek social interactions to accomplish shared goals, to enjoy shared interests, and to share information. Indeed, it is through social interactions that people come to learn about themselves, others, and the world around them.

The successful navigation of social interactions depends on a range of cognitive faculties. Chief amongst these, and the focus of the current dissertation, is the ability to achieve joint attention. Joint attention is defined as the simultaneous coordination of attention between a social partner and an object or event of interest (Bruner, 1974, 1995). Joint attention is typically achieved when one person *initiates* joint attention (IJA) by intentionally guiding their social partner to attend to an object or event in the immediate environment, and then the social partner *responds* to that joint attention bid (RJA). Individuals must then *evaluate* the achievement of joint attention (EAJA) to determine whether further communication is needed in order to share attention with their social partner (Tomasello, 1995).

Joint attention is considered to be an important precursor to the typical development of language and social cognition (Adamson, Bakeman, Deckner, & Romski, 2009; Baron-Cohen, 1995; Charman, 2003; Mundy, Sigman, & Kasari, 1990; Murray et al., 2008). Conversely, impaired joint attention is a characteristic of some types of atypical development, such as autism (American Psychiatric Association, 2013; Hill & Frith, 2003). The severity of joint attention impairment in autism is used to help define the degree of diagnostic severity, as well as predict the likely development of secondary impairments such as language delay (Charman, 2003; Dawson et al., 2004; Mundy et al., 1990).

While much is known about the relationship between joint attention and its impact on development, there has been little empirical investigation of the cognitive and neural

mechanisms underlying joint attention. This is largely due to the absence of ecologically valid experimental paradigms for the measurement of joint attention. In a recent review, Schilbach and colleagues (2013) outlined the challenges associated with achieving ecologically valid measures of social cognition, particularly in neuroscience research. Traditional approaches have relied on non-interactive or "offline" social cognition paradigms in which participants observe and respond to social stimuli (e.g., faces, and gaze cues) but do not "interact" with them. Schilbach and colleagues argue that offline paradigms cannot truly capture the cognitive and neural mechanisms of social cognition because the demands on these mechanisms are likely to be fundamentally different when engaged in "online" real-time social interactions. Instead, "second person" interactive paradigms should be employed when investigating elements of social cognition that are intrinsically interactive (Schilbach et al., 2013).

The second person approach is especially pertinent to the investigation of joint attention, which can only occur within social interactions. For this reason, joint attention paradigms must allow participants to interact with a social partner from a second person (i.e., you and I) perspective. However, it is challenging to develop a paradigm that simulates a real-life interaction and yet maintains experiment control. Accordingly, the first aim of this dissertation was to develop an ecologically valid interactive paradigm for the measurement of joint attention in behavioural and neuroimaging contexts (Chapter 2). The second aim was to apply this new paradigm to investigate the cognitive and neural mechanisms of joint attention in people with typical development (Chapter 3, 5, and 6) and in individuals with autism (Chapter 4).

In the remainder of this introductory chapter, I will provide an overview of the literature most relevant to these aims. I will start by reviewing the literature surrounding the ability to achieve joint attention. I will review the Parallel-Distributed Processing Model (PDPM) account of how the cognitive and neural mechanisms of RJA and IJA

develop (Mundy & Jarrold, 2010). Then I will review the interactive paradigms that have been developed to provide experimental measures of RJA and IJA in neuroimaging environments. Following this, I will review the literature concerning the neural mechanisms that support our ability to evaluate the achievement of joint attention (i.e., EAJA). Finally, I will highlight the outstanding questions that are addressed in the studies that comprise this dissertation.

Responding to and Initiating Joint Attention Bids (RJA and IJA)

Humans achieve joint attention with another person during social interactions by engaging in RJA or IJA behaviours. In children with typical development, RJA behaviours begin to emerge at approximately six months-of-age (Bakeman & Adamson, 1984; Scaife & Bruner, 1975) whilst IJA behaviours develop later at approximately 12 months-of-age (Bates, Benigni, Bretherton, Csmaioni, & Volterra, 1979). In children with autism, RJA behaviours start to emerge once cognitive development is equivalent to 30 to 36 months-of-age (Mundy, Sigman, & Kasari, 1994), whilst IJA behaviours can be impaired well into adolescence (e.g., Charman, 2003; Hobson & Hobson, 2007; Sigman & Ruskin, 1999). Individual variation in joint attention is a reliable predictor of social functioning (Lord et al., 2000; Stone et al., 1997) and language development in children with autism (Charman, 2003; Dawson et al., 2004). Thus, there is substantial evidence that RJA and IJA have different developmental trajectories in typical development and in autism.

The Parallel-Distributed Processing Model (PDPM)

The PDPM of joint attention provides an account for the developmental divergence of RJA and IJA in typical development (Mundy & Jarrold, 2010; Mundy & Newell, 2007; Mundy, Sullivan, & Mastergeorge, 2009). The PDPM builds on work by Posner and colleagues who suggest that RJA is supported by a posterior attention network in infancy, and IJA depends upon an anterior attention network that emerges later in development (e.g., Posner & Rothbart, 2007). The posterior attention network – comprising the

precuneus, posterior parietal cortex, occipital association cortex, intraparietal sulcus, and posterior superior temporal sulcus (pSTS) – is thought to underpin cognitive processes related to the processing of spatial information, executing rapid shifts in attention, gaze perception, and prioritising the processing of animate stimuli (e.g., eye gaze and gestures; see Figure 1). The anterior attention network – which comprises the frontal eye fields (including the superior colliculus pathway), anterior cingulate cortex (ACC), medial prefrontal cortex (mPFC) and orbitofrontal cortex (OFC) – is thought to support cognitive processes associated with the ability to represent another's perspective, supress automatic eye movements, and execute goal-directed shifts of attention to engage in rewarding experiences.
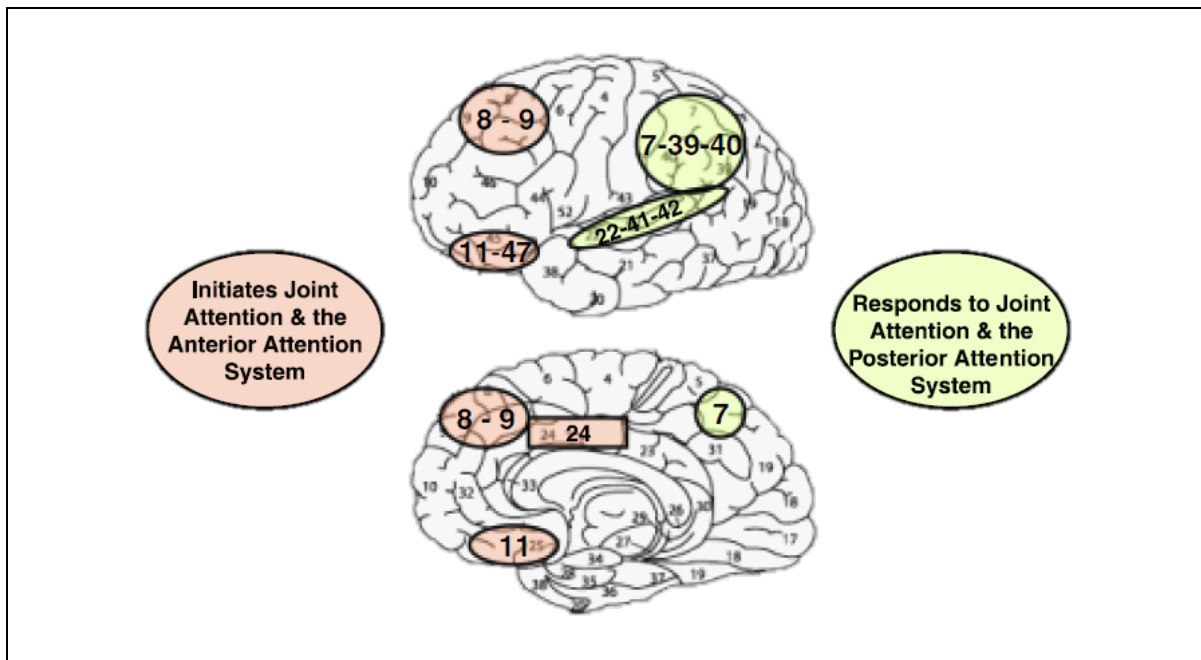
Figure 1. This figure, taken from Mundy and Newell (2007), depicts the posterior RJA network (yellow regions) and the anterior IJA network (orange regions) that are hypothesised to support joint attention behaviours. The numbers on the figure associate with Brodmann areas. RJA: 7 (posterior parietal association area); 22, 41, and 42 (superior temporal cortex); and 39 and 40 (parietal, temporal, and occipital association cortices). IJA: 8 (frontal eye fields), 9 (prefrontal association cortex), 24 (dorsal anterior cingulate), 11 and 47 (orbital prefrontal association cortex; Mundy and Newell, 2007, p. 271). This figure is reprinted with permission.

A critical claim made by the PDPM is that the posterior and anterior neural networks, which operate separately in early infancy, begin to integrate with development. This integration is associated with the emergence of complex joint attention behaviours during dynamic interactions, including the ability to simultaneously represent self- and other-oriented attentional perspectives (see Figure 2). Representing these perspectives in parallel allows individuals to align their behaviour to changes in their partner's perspective. For example, an individual may wait until they have their social partner's attention before attempting to initiate joint attention. This allows individuals to align their attentional perspectives with others and thus achieve joint attention.
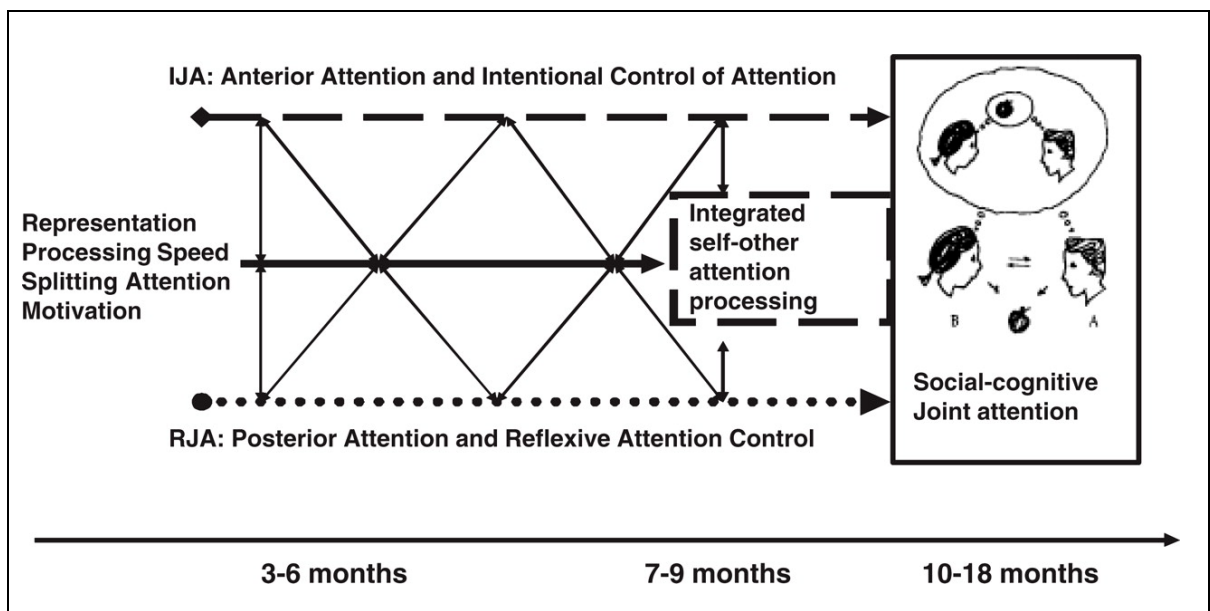
*Figure 2.*This figure, taken from Mundy and Newell (2007), depicts the developmental trajectory of RJA (dotted line) and IJA (dashed line). The solid lines represent the development of other cognitive mechanisms (e.g., executive control of attention, and mentalising processes). The development of these mechanisms have multidirectional influences on the development of both RJA and IJA. This results in the ability to represent self- and other-oriented perspectives of attention during social interactions. This figure is reprinted with permission.

The PDPM's neural model of joint attention has not yet been directly tested using experimental joint attention paradigms. This may be because it is difficult to design paradigms that capture the dynamic, spontaneous, and complex interactions involved in

joint attention without compromising experimental control. This challenge is exacerbated by the need to measure joint attention in neuroimaging environments that are typically built to test only one individual at a time.

Despite these challenges, several functional magnetic resonance imaging (fMRI) studies have attempted to develop well-controlled experimental measures of RJA and IJA that are both ecologically valid and executable within neuroimaging environments. Redcay et al. (2010; 2012) developed an interactive paradigm in which participants watched a computer screen whilst lying inside an MRI scanner. Their task was to collaborate with an experimenter (outside the scanner) via a live video feed in order to "catch" a mouse that was concealed behind one of four cheeses presented in each corner of the screen. On RJA trials, the experimenter was cued to the mouse's location by a tail protruding from one of the cheeses. This cue was not visible to the participant, who had to follow the experimenter's gaze to look at the correct location and hence catch the mouse. On IJA trials, participants were given the peripheral cue, and they were required to guide the experimenter to the mouse by looking at the appropriate location. In both RJA and IJA trials, the mouse could only be caught if joint attention was achieved at the correct location. This was monitored by a second experimenter outside the scanner.

In these studies, the RJA and IJA conditions were contrasted with a "solo attention" baseline condition that was identical to IJA trials except that the experimenter's eyes remained closed. Hence, the participant did not need to respond to or initiate joint attention in order to catch the mouse. Relative to this baseline, RJA trials were associated with increased activation in pSTS, dorsal mPFC, and posterior cingulate. In contrast, IJA trials were associated with increased activation in inferior frontal gyrus (IFG), bilateral anterior operculum, medial superior frontal gyrus, left middle frontal gyrus, right precentral gyrus, and inferior parietal lobe.

In another second person joint attention paradigm, Saito et al. (2010) used a

hyperscanning design in which two participants were scanned simultaneously as they interacted via live video feed. Each participant could see their partner's eyes at the top of the screen and two coloured circles below. On each trial, one participant was instructed to make an eye-movement in response to a colour change in one of the circles, and the other participant was instructed either to respond congruently to their partner's gaze and look at the same location or to respond incongruently and look at the other circle on the screen. The role of both participants changed throughout the task so that by the end of the experiment, each participant had completed four conditions of trials that each comprised different cues (i.e., partner's gaze or circle colour) and the outcome of joint attention (i.e., achieved or avoided). Thus, on each trial, participants (1) achieved joint attention intentionally by following a gaze cue, (2) failed to achieve joint attention by avoiding a gaze cue, (3) incidentally achieved joint attention by attending to a colour cue, or (4) incidentally avoided joint attention by attending to a colour cue. Saito et al. assessed the effect of following gaze to achieve joint attention (i.e., RJA) by comparing the neural response from all four conditions using the following interaction contrast, $[(1) - (2)] - [(3) - (4)]$. This was argued to isolate the neural activation associated with RJA. Unlike Redcay et al. (2010; 2012), Saito et al. (2010) failed to find mPFC activity related specifically to RJA trials, but they did report increased activation in the intraparietal sulcus for RJA.

In a third study by Schilbach et al., (2010), participants were told that they would engage in joint attention with an anthropomorphic virtual character whose eye-movements were controlled by a confederate outside the scanner using an eye-tracking device. In reality, a gaze-contingent algorithm (Wilms et al., 2010) was used to control the virtual character's eye-movements. When lying in an MRI scanner, participants were shown the virtual character's face presented in the middle of a screen. Three squares were positioned to left, right, and above the virtual character's face. On RJA trials (referred to as OTHER_JA by Schilbach et al., 2010, p. 2702), the virtual character averted his gaze to

one of the three squares presented on the screen. Participants were instructed to respond

congruently by looking at the same square as their partner. Baseline trials for the RJA

condition (referred to as OTHER_NOJA by Schilbach et al., 2010, p. 2706) were identical

to RJA trials except that participants were instructed to respond incongruently to their

partner's gaze by fixating on one of the two squares not being attended by their partner.

Subtracting brain activation in baseline control trials from RJA trials revealed differential

activity in the ventral mPFC. This region has previously been implicated in tasks involving

representing the mental states of others (Amodio & Frith, 2006). It has also been

associated with gaze processing in contexts where the perspective of the participant and the

"gazer" are aligned (EAJA, discussed below; Williams, Waiter, Perra, Perrett, & Whiten,

2005).

On IJA trials (referred to as SELF_JA by Schilbach et al., 2010, p. 2702),

participants were instructed to initiate joint attention by fixating one of the three squares.

The virtual character responded congruently by looking at the same location. This was

contrasted with a baseline condition that was identical to IJA trials except that the virtual

character responded incongruently by looking at a different square than the one fixated by

the participant. This contrast revealed activation in the ventral striatum, which has been

hypothesised to reflect the hedonic experience of evaluating the successful outcome of IJA

(Schilbach et al., 2010). This hypothesis is supported by the fact that the only difference

between IJA and baseline trials was whether the virtual character achieved or avoided joint

attention. Thus, the difference in neural activity between IJA and baseline trials related to

the evaluation of whether joint attention had been achieved (EAJA, discussed in the

following section) rather than the mechanisms involved in executing IJA behaviour.

These three ground-breaking studies represent an important step towards

investigating the neural correlates of joint attention and providing an empirical validation

of the PDPM. However, as I discuss in **Chapter 2**, they also raise a number of

methodological issues. For example, one methodological feature common to all the aforementioned paradigms is that participants were overtly instructed about their social role (i.e., responder or initiator) at the beginning of each trial or block. This departs from real social interactions whereby individuals must monitor the attention and intentions of their social partner in order to determine their social role and identify opportunities for RJA and IJA. Furthermore, informing participants of their social role makes the gaze behaviour of the participant's social partner predictable and unambiguous. This reduces the need for participants to interpret "ostensive" cues (e.g., eye contact) that are typically used to disambiguate gaze shifts that signal intentional bids for communication from those that do not convey any social meaning (Cary, 1978). Therefore, in the attempt to distil the complexity of social interactions within experimental protocols that provide critical control, previous studies may have failed to capture the attention and intention monitoring processes which pre-empt adaptive joint attention behaviours.

In an effort to further advance the measurement of joint attention at the level of the brain and behaviour, **Chapter 2** outlines a new second person joint attention paradigm that I have developed to overcome some of the persistent compromises that exist between experimental control and ecological validity. This "Catch-the-Burglar" paradigm simulates a social interaction that (1) elicits intentional goal-driven joint attention behaviours, (2) naturally informs participants of their social role without overt instruction, and (3) requires participants to monitor the attention and intentions of their social partner throughout the interaction in order to correctly interpret the communicative information conveyed by gaze. This paradigm also includes non-social baseline conditions that are identical to RJA and IJA conditions except for RJA- and IJA-related task demands (e.g., they control for task complexity, number of eye movements elicited, and non-social attentional demands). Subtracting the behavioural or neural responses to these baseline control trials from responses to RJA or IJA trials can be used to elucidate the cognitive and neural

mechanisms that are specific to RJA and IJA.

In **Chapter 3**, I apply the Catch-the-Burglar paradigm in an fMRI experiment to test some of the PDPM's claims. As mentioned earlier, the PDPM suggests that RJA is supported by a posterior-parietal network while IJA is supported by an anterior network. These neural networks, which are though to be discrete in early development, are argued to integrate as development progresses to support social cognition and joint attention behaviours during social interactions (Mundy & Jarrold, 2010; Mundy & Newell, 2007; Mundy, Sullivan, & Mastergeorge, 2009). Thus, according to the PDPM, in adulthood RJA and IJA should be supported by a common frontotemporoparietal network comprising the regions identified in the PDPM. The use of properly matched baseline conditions in the Catch-the-Burglar paradigm means that the neural activation specifically associated with RJA and IJA can be isolated and compared. Using this paradigm, I examine the conjunction of RJA and IJA effects to identify the neural correlates common to these joint attention functions.

In **Chapter 4**, I use the same paradigm to investigate whether adults with high-functioning autism (HFA) demonstrate impairments in RJA and IJA behaviours. The tightly-matched non-social baseline conditions in the Catch-the-Burglar paradigm allow me to elucidate whether joint attention impairments in autism are specific to processes of social cognition, or whether they can be explained by impairments in non-social cognitive functions which impede performance on the interactive task (e.g., attention, and executive control). A second person approach is especially important in supporting our understanding of joint attention difficulties in autism given that impairments of social cognition are likely to be rooted in processes that are contingent on interactive dynamics (Becchio, Sartori, & Castiello, 2010; Schilbach et al., 2013). As such, we cannot understand these impairments by investigating offline social cognition. This may help explain why the consistent findings of joint attention impairments in naturalistic studies of

joint attention behaviour (e.g., Hobson & Hobson, 2007; Mundy et al., 1994) are not matched by results of gaze-orienting impairments in autism using non-interactive gaze-cueing paradigms (see Nation & Penny, 2008 for a review). I discuss these issues in my review of experimental methodologies for the measurement of RJA and IJA behaviour in **Chapter 2**.

Evaluating the Achievement of Joint Attention (EAJA)

Theoretical and empirical accounts of joint attention have predominantly focused on behaviours that result in the achievement of joint attention experiences (i.e., RJA and IJA). However, another important process is evaluating whether joint attention has been achieved (EAJA). This may have received less attention in the scientific literature because EAJA is not characterised by overt behaviour. Nevertheless, the ability to determine whether joint attention has been achieved is important in coordinating social interactions (Tomasello, 1995). This is particularly the case after we have attempted to initiate joint attention, as it allows us to determine whether we have successfully guided our social partner's attention, or whether we need to try again.

Emery (2000) has argued that the mutual awareness of joint attention between two individuals gives rise to a phenomenon separate to joint attention that he calls "shared attention". However, in this dissertation I consider EAJA to be a sub-process of "joint attention" for two reasons. First, the majority of the relevant experimental literature considers EAJA to be a component of joint attention (cf. Pfeiffer, Vogeley, & Schilbach, 2013). Second, unlike Emery's definition of shared attention, EAJA does not depend on the mutual awareness of whether joint attention has been achieved. Instead, it refers to a single individual's evaluation or "awareness" of whether joint attention has been achieved.

Regardless of whether EAJA is considered to be an aspect of joint attention or shared attention, it is measured in the same way in experimental paradigms that test the ability of a single individual (i.e., a participant) to assess the achievement of a joint

attention with an experimenter or avatar. For example, as noted earlier, Schilbach et al. (2010) conducted a study that effectively identified the neural correlates of EAJA. To recap, participants were instructed to guide a virtual partner to a location on a computer screen. In some trials their partner responded congruently to achieve joint attention, and in other trials their partner responded incongruently and failed to achieve joint attention. Congruent responses were associated with greater activation of the ventral striatum than incongruent responses. The ventral striatum has been consistently recruited during reward processing paradigms (McClure, York, & Montague, 2004). Accordingly, Schilbach et al. argued that the activation of the ventral striatum in their study reflected the hedonic experience associated with achieving joint attention (i.e., EAJA).

Two further fMRI studies have investigated the neural correlates of EAJA. Both used paradigms in which participants were aware that they were not interacting with a real person. Williams et al. (2005) required participants to orient to a red dot in the bottom panel of a computer screen. A video of a man played in the top panel of the screen. In each trial, the man oriented his head to look at the red dot (simulating an experience of "joint attention") or he looked at the other side of the screen to avoid joint attention. Compared to trials that did not result in joint attention experiences, achieving joint attention was associated with increased activation in the ventral mPFC. The involvement of the mPFC in EAJA again aligns with previous findings associating this region with representing the mental states and perspectives of others (Amodio & Frith, 2006).

In a more recent study by Gordon, Eilbott, Feldman, Pelphrey, & Vander Wyk (2013), participants were presented with a pre-recorded video still of "Sally" who was flanked by two identical silhouettes. Participants guided Sally by fixating on one of the silhouettes, whereupon Sally turned her head to the fixated silhouette (congruent response) or the non-fixated silhouette (incongruent response). Compared to incongruent responses, viewing congruent (joint attention) responses resulted in greater activation of ACC, right

fusiform gyrus, right amygdala, striatum and parahippocampus. The authors interpreted these regions as supporting the experience of social reward when joint attention was achieved. This pattern of activation differs from Williams et al.'s (2005) findings despite the use of very similar paradigms. I discuss possible reasons for these different outcomes in detail in **Chapter 2**.

Whilst the location of the neural correlates of EAJA have been investigated by several fMRI studies, to date, no study has examined the timing of the neural processes associated with EAJA. The studies reported in **Chapter 5** and **Chapter 6** attempt to address this gap in the literature. These studies were motivated by an event-related potential (ERP) study reported by Carrick, Thompson, Epling, and Puce (2007). Carrick et al. presented participants with an array of three photographic faces arranged in a horizontal line: two flanker faces on either side of a central face. The flanker faces gazed in the same direction (left or right) for the duration of each trial. The gaze of the central face, which was initially directed toward the participant, shifted in one of three directions in each trial to establish one of three social contexts. In a "group" condition, the central face's gaze was shifted in the same direction as the flanker faces to establish an instance of shared attention. In a "mutual" condition, the central face's gaze was shifted in the opposite direction to the flanker faces, establishing an instance of mutual gaze with one of the flanker faces. And in an "avoid" condition, the central face's gaze was shifted vertically upward, thus avoiding both mutual gaze and shared attention with either of the flanker faces. ERPs time-locked to the onset of the central face's gaze shift were measured over occipitotemproal and centro-parietal sites on the scalp. While gaze shifts evoked an occipitotemporal N170 response, this was not modulated by condition, which is consistent with the idea that the N170 is an index of the early perceptual encoding of faces and eye gaze (Itier & Taylor, 2004). In contrast, gaze shifts by the central face evoked centro-parietal P350 and P500 ERPs that were modulated by condition. Specifically, gaze shifts

reflecting instances of shared attention or mutual gaze resulted in an earlier P350 ERP and smaller P500 ERP relative to gaze shifts reflecting an avoidant gaze scenario. This suggested that the neural processes associated with evaluating the social significance of gaze occurs within 350 and 500 ms of observing a gaze shift.

While Carrick et al.'s (2007) findings are encouraging, they are somewhat limited by the ecological validity of their paradigm. Specifically, this paradigm did not allow participants to interact with the central or flanker faces. Thus, Carrick et al.'s paradigm cannot be used to elucidate the neural processes associated with EAJA during a social interaction. To address this issue, I conducted an ERP study (see **Chapter 5)** that used an adapted version of the Catch-the-Burglar paradigm. In this modified paradigm, participants were told that their virtual partner was a "prison guard" responsible for patrolling the inside of the prison. Participants were also informed that the guard's job was to "lock down" any "breached" exits. Participants were instructed that their role was to patrol the outside of the prison. They were told that on each trial a convicted burglar would try and escape from the prison compound. The participant was required to fixate upon the appropriate exit to show their partner (i.e., the guard) where the burglar was trying to escape from, and hence which exit to lock down. Participants were then required to fixate back on their partner's face to determine if he responded appropriately.

To ensure that participants observed their partner's response, the virtual character was programmed to only respond once the participant had fixated back on his face. Participants were informed that their partner might sometimes be distracted by scuffles within the prison. Accordingly, on 50% of trials, the virtual character responded congruently to the participant's joint attention bid, and on 50% of trials he responded incongruently by gazing towards an incorrect location. The contrast between ERP responses on congruent and incongruent trials revealed the time course of neural processes associated with EAJA.

In a final study, outlined in **Chapter 6**, I use the same Catch-the-Burglar paradigm with ERPs to address a slightly different issue. That is, does it matter whether people believe the virtual character is a human-controlled avatar, or a computer-programmed agent? I test this by comparing the ERPs associated with EAJA when participants believed they were interacting with another person, with the ERPs elicited in a second group of participants who completed the same task, but believed their virtual partner was being controlled by a computer program.

Summary and Overview

Joint attention plays a critical role in social interactions and in cognitive and linguistic development. However, given the technical and methodological difficulties involved in measuring joint attention in behavioural and neuroimaging contexts, we currently have a very limited understanding of the cognitive and neural mechanisms that underlie RJA, IJA, and EAJA. The research contained in this dissertation addresses two overarching aims. The first is to develop an ecologically valid and experimentally controlled paradigm for the measurement of joint attention in behavioural and neuroimaging contexts. Following a critical review of existing methodologies, I present a new ecologically valid paradigm (**Chapter 2**). The second aim is to apply this new paradigm to investigate the cognitive and neural mechanisms of joint attention in people with typical development and autism spectrum disorders. Accordingly, I use this paradigm in an fMRI study that investigates the neural correlates of RJA and IJA in typically developing adults (**Chapter 3**). Then, I use the same paradigm in an eye-tracking study to investigate joint attention impairments in autism, using a range of eye tracking measures to determine whether individuals with HFA differ to peers with typical development regarding the accuracy and speed with which they are able to execute RJA and IJA behaviours (**Chapter 4**). Next, I modify my paradigm and apply it in an ERP study to investigate the time course of neural processes associated with EAJA (**Chapter 5**). I then

use the same modified paradigm to investigate whether the neural processes associated

with EAJA are influenced by participants' beliefs about whether their virtual partner is

being controlled by a real human or a computer program (**Chapter 6**). Finally, I provide a

general discussion of all the studies presented in this dissertation (**Chapter 7**).

References

Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. A. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of Autism & Developmental Disorders, 39*(1), 84-96. doi: 10.1007/s10803-008-0601-7

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5 ed.). Washington, DC

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268-277. doi: 10.1038/nrn1884

Bakeman, R., & Adamson, L. B. (1984). *Child Development, 55*(4), 1278. doi: 10.1111/1467-8624.ep7302943

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

Bates, E., Benigni, L., Bretherton, I., Csmaioni, L., & Volterra, V. (1979). *The emergence of symbols: cognition and communication in infancy.* New York: Academic Press.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497-529. doi: 10.1037/0033-2909.117.3.497

Becchio, C., Sartori, L., & Castiello, U. (2010). Toward You: The Social Side of Actions. *Current Directions in Psychological Science, 19*(3), 183-188. doi: 10.1177/0963721410370131

Bruner, J. S. (1974). From communication to language, A psychological perspective. *Cognition, 3*(3), 255-287. doi: 10.1016/0010-0277(74)90012-2

Bruner, J. S. (1995). From joint attention to the meeting of minds: an introduction. In C. Moore & P. J. Dunham (Eds.), *Joint attention: its origins and role in development.* (pp. 1-14). Hillsdale, NJ: Lawrence Erlbaum Asscoiates.

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. doi: 10.1097/WNR.0b013e3280ebb44b

Cary, M. S. (1978). The Role of Gaze in the Initiation of Conversation. *Social Psychology, 41*(3), 269-271. doi: 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. doi: 10.1098/rstb.2002.1199

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. *Developmental Psychology, 40*(2), 271-283.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*(6), 581-604. doi:10.1016/S0149-7634(00)00025-7

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. doi: 10.1080/17470919.2013.832374

Hill, E. L., & Frith, U. (2003). Understanding autism: insights from mind and brain. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1430), 281-289. doi: 10.1098/rstb.2002.1209

Hobson, J., & Hobson, P. (2007). Identification: The missing link between joint attention and imitation? *Development and Psychopathology, 19*(02), 411-431. doi: doi:10.1017/S0954579407070204

Itier, R. J., & Taylor, M. J. (2004). Face recognition memory and configural processing: a

    developmental ERP study using upright, inverted, and contrast-reversed faces.

    Journal of Cognitive Neuroscience, 16(3), 487-502. doi:

    10.1162/089892904322926818

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. *Journal of Autism and Developmental Disorders, 30*(3), 205–223. doi:

    10.1023/A:1005592401947

McClure, S. M., York, M. K., & Montague, P. R. (2004). The Neural Substrates of Reward

    Processing in Humans: The Modern Role of fMRI. *The Neuroscientist, 10*(3), 260-

    268. doi: 10.1177/1073858404263526

Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social

    cognition. *Neural Networks, 23*(8), 985-997. doi: 10.1016/j.neunet.2010.08.009

Mundy, P., & Newell, L. (2007). Attention, joint attention and social cognition. *Current

    Directions in Psychological Science, 16*, 269-274. doi: 10.1111/j.1467-

    8721.2007.00518.x

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and

    language development in autistic children. *Journal of Autism and Developmental

    Disorders, 20*(1), 115-128. doi: 10.1007/bf02206861

Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and

    symptom presentation in autism. *Development and Psychopathology, 6*(03), 389-

    401. doi: doi:10.1017/S0954579400006003

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-

    processing model of joint attention, social cognition and autism. *Autism Research,

    2*(1), 2-21. doi: 10.1002/aur.61

Murray, D. S., Creaghead, N. A., Manning-Courtney, P., Shear, P. K., Bean, J., & Prendeville, J. A. (2008). The relationship between joint attention and language in children with autism spectrum disorders. *Focus on Autism & Other Developmental Disabilities, 23*(1), 5-14. doi: 10.1177/1088357607311443

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? Is it automatic? Is it social? *Development and Psychopathology, 20*(01), 79-97. doi: doi:10.1017/S0954579408000047

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience and Biobehavioural Reviews, 37*(10), 2516-2528. doi:10.1016/j.neubiorev.2013.07.017

Posner, M. I., & Rothbart, M. (2007). *Educating the human brain.* Washington, DC: American Psychological Association.

Redcay, E., Dodell-Feder, D., Mavros, P. L., Kleiner, A. M., Pearrow, M. J., Triantafyllou, C., Gabrieli, J. D., & Saxe, R. (2012). brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. *Human Brain Mapping, 34,* 2511-2523. doi: 10.1002/hbm.22086.

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. doi: 10.1016/j.neuroimage.2010.01.052

Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., . . . Sadato, N. (2010). "Stay Tuned": inter-individual neural synchronization during mutual gaze and joint attention. *Frontiers in Integrative Neuroscience, 4*, 127. doi: 10.3389/fnint.2010.00127

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature, 253*(5489), 265-266.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393-414. doi: 10.1017/S0140525X12000660.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah N. J., Fink, G. R., & Vogeley, K. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience, 22*(12), 2702-2715. doi: 10.1162/jocn.2009.21401.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome and developmental delays. *Monographs of the Society for Research in Child Development, 64*(1), 1-114. doi: 10.1111/1540-5834.00002

Stone, W. L., Ousley, O. Y., & Littleford, C. D. (1997). Motor imitation in young children with autism: what's the object? *Journal of Abnormal Child Psychology, 25*, 475-485. doi: 10.1023/A:1022685731726

Tomasello, M. (1995). Joint Attention as Social Cognition. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development*. Hillsdale: Lawrence Erlbaum Associates

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. doi: 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. doi: 10.1093/scan/nsq024

Chapter 2

The Measurement of Joint Attention:

A Review, Recommendations, and New Paradigm

Nathan **Caruana**[123], Alexandra **Woolgar**[123], Genevieve **McArthur**[12], & Jon **Brock**[124]

[1] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[2] ARC Centre of Excellence in Cognition and its Disorders

[3] Perception in Action Research Centre, Sydney

[4] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

Abstract

The ability to achieve joint attention with others is a cornerstone of everyday social interactions and supports the development of language and social learning skills. Impairments in joint attention are also characteristic of autism spectrum disorders. Despite the importance of joint attention in typical development and in autism, little is known about its underlying cognitive and neural mechanisms. A key challenge is the limited availability of paradigms that can provide experimentally controlled and ecologically valid measures of joint attention. Given that joint attention can only be experienced during social interactions, a "second person" approach is needed whereby participants can engage in genuine (or seemingly genuine) real-time social interactions with others (Schilbach et al., 2013). In this paper we review the studies that have pioneered second person paradigms for the measurement of joint attention. We highlight the outstanding methodological challenges associated with the measurement of joint attention in behavioural and neuroimaging experiments, and provide recommendations for overcoming these challenges. We also present a new paradigm that builds on the innovations of previous second person approaches to provide a controlled and ecologically valid method for the experimental investigation of joint attention.

The Measurement of Joint Attention:

A Review, Recommendations, and New Paradigm

Introduction

Joint attention refers to the ability to simultaneously coordinate attention between a social partner and an object or event of interest (Bruner, 1974). It is a communicative act. One person initiates joint attention (IJA) by intentionally directing their social partner to a particular location via eye gaze, head turns, gesture (e.g., pointing), or vocalization. The other person must recognise these behaviours as having communicative intent, and respond to the joint attention bid (RJA) by attending to the same location. Finally, at least one individual must determine whether they have been successful in achieving joint attention (Tomasello, 1995). We refer to this last ability as evaluating the achievement of joint attention (EAJA).

The ability to achieve joint attention is considered crucial for coordinating joint actions, language development, and social learning (Adamson, Bakeman, Deckner, & Romski, 2009; Baron-Cohen, 1995; Charman, 2003; Mundy, Sigman, & Kasari, 1990; Murray et al., 2008; Tomasello, 1995). For instance, if a parent is describing an object to an infant whilst directing the infant's attention to that object (i.e., initiating a joint attention bid), and the infant attends to that object (i.e., responding to a joint attention bid), the infant forms associations between the phonological, visual, and semantic representations of that word (Baldwin, 2014). Conversely, delay in the development of joint attention is one of the earliest visible symptoms of the social and communication impairments that characterise autism spectrum disorders (Frith, 2003). Amongst children with autism, individual variation in joint attention is a reliable predictor of social functioning (Lord et al., 2000; Stone, Ousley, & Littleford, 1997) and language development (Charman, 2003; Dawson et al., 2004; Mundy et al., 1990).

Current Approaches for Measuring Joint Attention

Despite its critical role in both typical and atypical development, little is known about the cognitive and neural mechanisms that support joint attention. Whilst cognitive (Baron-Cohen, 1995) and neural (Mundy et al., 1990) models of joint attention have been proposed, hypothesising how joint attention behaviours emerge throughout development, these models are largely descriptive, and have not been supported directly by empirical evidence. This is due, in part, to the difficulty in creating adequate experimental measures of joint attention.

According to the "second person" approach outlined by Schilbach and colleagues (2013), social behaviours should be measured whilst participants are immersed in a social interaction. This is particularly important for measuring joint attention, partly because joint attention depends upon reciprocal and ongoing social interactions, and partly because joint attention behaviour is greater than (or at least different to) the combined behaviours of each individual acting alone (Hobson, 2008). An ideal experimental measure of joint attention would capture the full complexity of joint attention whilst maintaining experimental control (Pfeiffer, Vogeley, & Schilbach, 2013; Schilbach et al., 2013).

Various paradigms have been used to measure three key components of joint attention: RJA, IJA, and EAJA. These paradigms rely almost exclusively on gaze-based interactions (see Pfeiffer et al, 2013, for a review). This is a sensible starting point because (1) communicative gaze is the first modality in which joint attention behaviours develop (Baron-Cohen, 1995; Mundy & Newell, 2007), (2) gaze provides an objective measure of an individual's locus of attention (Kowler, Anderson, Dosher, & Blaser, 1995), and (3) gaze behaviour is the best predictor of abnormal joint attention development, which is a feature of autism spectrum disorders amongst infants (Lord et al., 2000; Stone et al., 1997). Thus, in this review, we will focus on joint attention paradigms that have employed gaze-based interactions. We will not discuss the outcomes of all these studies as they have

already been comprehensively reviewed elsewhere (Pfeiffer et al. 2013). Instead, we focus upon the strengths of these paradigms and describe a new joint attention paradigm that merges these strengths within a second person approach. We also discuss the critical issues affecting the measurement of joint attention, and offer a number of recommendations for tackling the trade-off between ecological validity and experimental control when designing a joint attention paradigm.

Responding to Joint Attention Bids

According to the Early Social-Communication Scale (Mundy et al., 2003; Seibert, Hogan, & Mundy, 1982), RJA comprises "lower level behaviours" such as eye contact, gaze alternating, and following proximal pointing gestures, as well as "higher order behaviours" such as "following line of regard" beyond an object to follow a social partner's gaze or pointing gesture. Infants begin to display lower level RJA behaviours at approximately six months of age when they reflexively follow the gaze of others around them (Mundy, Sigman, & Kasari, 1994). In later development, RJA becomes increasingly complex as individuals learn to discriminate between gaze shifts that do and do not convey social meaning (Senju & Johnson, 2009).

The majority of experimental RJA studies have focused on measuring lower level gaze orienting using adaptations of the Posner-cueing paradigm (Posner, 1980). These studies typically ask participants to detect a visual target that is presented on the left or right side of a display. The target is preceded by gaze cues (i.e., a pair of eyes looking left or right) directing attention either towards (i.e., valid cue) or away from (i.e., invalid cue) the target location. This produces a "validity effect" that is characterised by a faster detection of targets preceded by a valid cue than invalid cue. Gaze cues give rise to validity effects even if participants are told that the gaze cue is irrelevant to the task and does not predict the target's location (Friesen & Kingstone, 1998). This suggests that eye gaze leads to the "reflexive" orienting of attention. The underlying assumption made by gaze-cueing

studies is that the reflexive orienting to gaze – as measured by the validity effect – is an adaptive behavioural response that supports RJA in typical development.

In terms of atypical development, it has been suggested that people with autism have RJA impairments stemming from a reduced sensitivity to gaze cues (Baron-Cohen, 1995). However, gaze-cueing studies have produced inconsistent evidence for this claim (see Nation & Penny, 2008 for a review). Some studies report reduced or absent validity effects in individuals with autism (e.g., Johnson et al., 2005; Ristic et al., 2005) whilst others do not (e.g., Chawarska, Klin, & Volkmar, 2003; Okada, Sato, Murai, Kubota, & Toichi, 2003; Vlamings, Stauder, van Son, Mottron, & Laurent, 2005). These contradictory findings might occur because social competencies are heterogeneous across individuals with autism (Dawson et al., 2002; Happe, Ronald, & Plomin, 2006; Volkmar, Cohen, Bregman, Hooks, & Stevenson, 1989). They might also be explained by methodological differences between studies such as: (1) the use of different aesthetic properties of gaze cues (e.g., photographs of real faces (Hood, Willen, & Driver, 1998), animated virtual faces (Pelphrey, Singerman, Allison, & McCarthy, 2003), and schematic faces (Friesen & Kingstone, 1998)); (2) the use of static versus dynamic gaze shifts; (3) the ratio of trials with valid and invalid cues; and (4) the cue-target stimulus onset asynchrony (see Frischen, Bayliss, & Tipper, 2007; Nation & Penny, 2008 for reviews).

Another factor that complicates our understanding of RJA in individuals with both typical and atypical development (i.e., autism), is the specificity of gaze-cueing tasks. Unlike naturalistic studies of RJA, which have consistently reported evidence for RJA impairments in autism (Charman et al., 1997; Dawson et al., 2004; Loveland & Landry, 1986; Mundy et al., 1990; Osterling, Dawson, & Munson, 2002; Wong & Kasari, 2012), experimental gaze-cueing tasks are not designed to tax the full range of processes involved in RJA. Instead, they are designed to tax one facet of RJA, which is the ability to reflexively orient attention in response to a gaze cue. This specific ability differs from real

social interactions, since it would be inappropriate to orient to every gaze shift made by a social partner. Rather, individuals need to monitor their social partner's gaze shifts to determine if they are intended to be communicative (and hence present an opportunity for RJA) or whether they have no social meaning (Senju & Johnson, 2009). This judgement is typically achieved by interpreting "ostensive" cues conveyed by one's social partner (e.g., eye contact; Cary, 1978; Tylén, Allen, Hunter, & Roepstorff, 2012). Thus, it is important that RJA paradigms tax the ability to monitor a social partner's intention to communicate.

A series of studies by Leekam and colleagues came closer to measuring ecologically-valid intention monitoring processes by observing RJA behaviours during genuine, yet structured, gaze-following interactions (Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997; Leekam, Lopez, & Moore, 2000). Leekam et al. (2000) developed a structured gaze-following task in which an adult experimenter sat at a table across from an infant participant. Two boxes were positioned on the table, one to the left and one to the right of the experimenter. On each trial, the adult established eye contact with the infant and then looked at one of the boxes. When the infant demonstrated RJA, the appropriate box was opened to reveal a toy and flickering lights. The time taken for the infant to follow the adult's joint attention bid was measured offline from video recordings of the interaction.

Leekam et al.'s (2000) gaze-following paradigm differed from previous gaze-cueing paradigms by providing a direct measure of naturalistic gaze-following behaviour. Specifically, it allowed the adult social partner to establish eye contact with the infant participant to communicate that their gaze shift was an intentional bid for joint attention. However, an unfortunate consequence of using this genuine interaction was the loss of experimental control. It was not possible to control variation in the adult's behaviour across trials or between participants, and the offline coding of videos provided a relatively crude measure of performance.

In an attempt to remedy these limitations, several computer-based eye tracking tasks have been developed to simulate structured gaze-following interactions whilst providing better control over the stimulus, and greater spatial and temporal precision in the measurement of RJA performance (Gredebäck, Theuring, Hauf, & Kenward, 2008; Senju & Csibra, 2008; von Hofsten, Dahlström, & Fredriksson, 2005). Navab, Gillespie-Lynch, Johnson, Sigman and Hutman (2011) simulated an interactive context based on Leekam et al.'s (2000) task. On each trial, an animated sequence was used to encourage infants to fixate the centre of the screen. This was followed by a pre-recorded video of an actor sitting at a table with coloured blocks in front of her to the left and right. Initially, the actor looked down at the centre of the table. Then she directed her gaze to the camera (to simulate eye contact with the infant), smiled, and said "hello there". She then initiated joint attention by gazing toward one of the blocks. The infant was given 4000 ms to respond by fixating the block attended by the actor.

Navab et al. (2011) used this paradigm in a sample of 18-month-old infants who had siblings with autism. They found that the percentage of accurate RJA gaze shifts was correlated with scores on the Early Social-Communication Scale. This demonstrated the task's construct validity as a measure of adaptive gaze-following behaviour during an everyday social interaction. However, the degree of ecological validity was limited by the use of pre-recorded videos. In particular, the pre-recorded actor in this paradigm initiated joint attention regardless of whether the infant was attending to the video or not. This contrasts with real social interactions in which an adult would typically wait to establish eye contact with an infant before initiating joint attention. This lack of reciprocity between the adult and infant diminished the relevance of the actor's communicative behaviours (e.g., eye contact and social greeting). Consequently, participants may have missed opportunities for RJA due to a lack of engagement rather than a reduced ability to produce RJA behaviours. Indeed, Navab et al. reported that their findings were consistent with

previous studies demonstrating reduced RJA behaviours when individuals responded to pre-recorded stimuli rather than real interactions (Gredebäck, Fikke, & Melinder, 2010). Thus, whilst Navab et al.'s paradigm provided superior experimental control compared to previous joint attention paradigms, it lacked an interactive interface that supported (or simulated) a truly reciprocal interaction between a participant and their social partner.
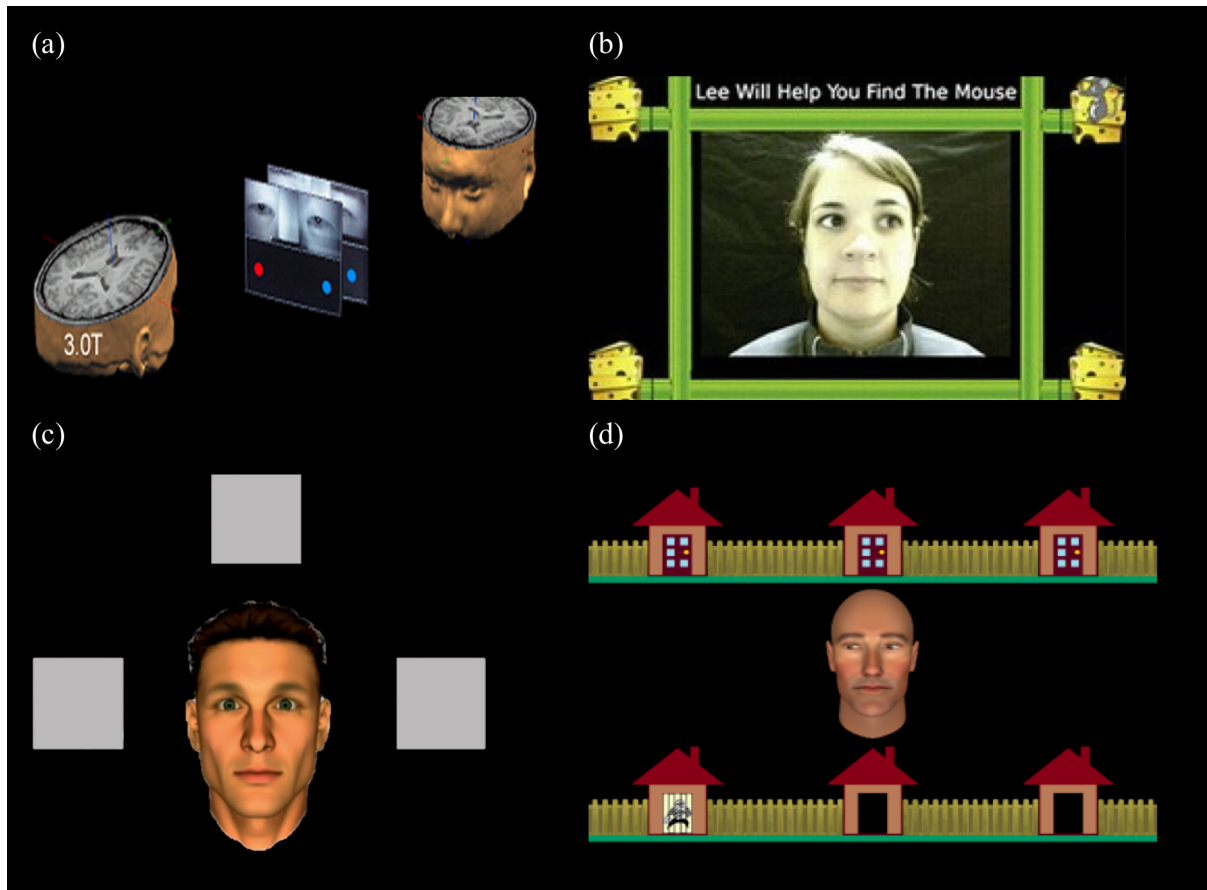
Figure 1. Still frames from interactive paradigms measuring RJA behaviour in fMRI studies. (a) Saito et al.'s (2010) hyperscanning paradigm. (b) Redcay et al.'s (2010; 2012) cooperative video interaction paradigm (taken from Redcay et al., 2010). (c) Schilbach et al.'s virtual reality paradigm (taken from Wilms et al., 2010). (d) Caruana, Brock & Woolgar's (2015) cooperative virtual reality paradigm. All figures reprinted with permissions.

In a response to this limitation, three functional magnetic resonance imaging (fMRI) studies developed computer interfaces that enabled truly reciprocal interactions. Saito et al. (2010) used a "hyperscanning" design in which two participants were scanned simultaneously in separate MRI scanners as they interacted via live video feed (see Figure 1a). Each participant could see their partner's eyes at the top of the screen. At the bottom of the screen, there were two coloured circles. On each trial, one of the participants was instructed to look for the circle that changed colour, while the other participant was instructed to respond to their partner's gaze and look at the circle in the same location. The role of each participant varied on each trial depending on the instruction provided.

Redcay et al. (2012) employed a similar task except that only one participant was tested in the MRI scanner. Their social partner was an experimenter ("Lee") who was presented via a live video feed. The participant was told that they had to help Lee catch a mouse that was concealed behind one of four cheeses that were located in each corner of the screen (see Figure 1b). On RJA trials, Lee was cued to the target location by a tail protruding from one of the cheeses. The participant was required to follow Lee's gaze in order to "catch" the mouse.

Schilbach et al. (2010) used a similar RJA paradigm to Redcay et al. (2012) except that participants interacted with an anthropomorphic virtual character rather than an experimenter (Wilms et al., 2010). Participants were told that a confederate outside the scanner controlled the virtual character's gaze via an eye-tracking device. In reality, the virtual character was controlled by an algorithm that was contingent on a participant's own eye movements (see Figure 1c). This gaze-contingent algorithm supported reciprocity in the interaction whilst maintaining control over the virtual partner's behaviour. On RJA trials (referred to as OTHER_JA, Schilbach et al., p. 2703), the virtual character averted his gaze towards one of three squares and participants were required to respond congruently by following his gaze.

These three fMRI studies used similar tasks that ostensibly measured similar cognitive processes (i.e., RJA). It is surprising, therefore, that they identified different patterns of brain activity related to RJA behaviour. Redcay et al. (2012) and Schilbach et al. (2010) identified activity in the medial prefrontal cortex (mPFC), a region implicated in mentalising (i.e., thinking about other people's thoughts; Amodio & Frith, 2006). However, Redcay et al. noted additional activity in posterior superior temporal sulcus and temporoparietal junction, with reduced activation among adults with autism. Saito et al. (2010) failed to find mPFC activity, but reported increased activation in the intraparietal sulcus.

While the discrepant findings across these three fMRI studies might be explained by differences in stimuli or social partners, it seems more plausible that they stemmed from major differences in the baseline conditions used to isolate RJA activation. fMRI analysis involves subtraction of changes in the blood-oxygen-level dependent (BOLD) response in one condition from changes in the BOLD response in another condition. This subtraction method relies on the assumption of pure insertion, which is the idea that two conditions (e.g., test and baseline conditions) only differ with respect to the cognitive process of interest (e.g., RJA). Thus, to isolate brain activation associated with RJA, it is necessary to compare the brain activation measured during RJA trials to brain activation measured during baseline trials that are matched to the RJA trials for all task demands except those specific to RJA (e.g., attentional processes, cognitive load, and eye movements).

Schilbach et al. (2010) and Saito et al. (2010) employed baseline trials in which participants were instructed to actively avoid joint attention by looking in the opposite direction to their partner. This is problematic for two reasons. First, the baseline condition required participants to make a saccade away from the cued location in order to "not follow" their partner's gaze. Relative to the RJA condition, this placed additional demands on attention and oculomotor control (cf. Pfeiffer et al., 2013). Second, both the RJA

37

condition and the baseline condition involved processing the gaze of the social partner and determining the location of the social partner's attention. Thus, rather than isolating activation associated with gaze processing, contrasting RJA and baseline trials removed activation related to gaze processing.

Redcay et al. avoided this problem by comparing RJA with a "solo attention" baseline condition in which a participant searched for a mouse while their partner's eyes remained closed. However, like Saito et al. (2010) and Schilbach et al. (2010), the baseline condition introduced additional task demands relative to the RJA condition. Specifically, on RJA trials, participants responded to gaze cues presented in the centre of the screen, whilst on baseline trials they engaged in a visual search for the mouse guided by a peripheral cue (i.e., a mouse tail). This meant that the baseline task placed more demands on eye movements and different demands on attention. Thus, the observed differences in activation between the baseline and RJA conditions employed by Schilbach et al. (2010), Saito et al. (2010), and Redcay et al. (2012) could, in part, represent differences in attention and oculomotor processes rather than RJA *per se*.

As well as using appropriate baseline conditions, it is important that studies of RJA capture the attention and intention monitoring processes involved in real-life RJA behaviour. In the three fMRI studies outlined above, the eye movements made by the participant's social partner were unambiguously communicative. Participants were also informed of their role as "responder" at the start of each block or trial, and thus knew that they had to follow every eye-movement made by their partner. However, in real social situations, individuals must determine which eye movements to follow and which to ignore (Cary, 1978; Tomasello, 1995). Thus, these interactive paradigms, despite their superior ecological validity relative to gaze-following and gaze-cueing paradigms, may not have fully captured the intention monitoring processes that are typically engaged in dyadic gaze-based interactions.

In sum, when measuring RJA, it is important to use paradigms that (1) simulate realistically complex social interactions that capture the intention monitoring processes integral to RJA behaviour, (2) allow participants to engage with a social partner whose behaviour is subject to experimental control, while supporting the experience of a reciprocal interaction, and (3) employ baseline conditions that are closely matched to the non-social task demands of the RJA condition. To this end, we developed a new paradigm – the "Catch-the-Burglar" paradigm - that combines three major strengths of RJA studies outlined above with several improvements (Chapters 3 & 4; Caruana, Brock & Woolgar, 2015). First, we adopted a similar approach to Schilbach et al. (2010) in asking participants to interact with a virtual character whom they believe to be controlled by a real person ("Alan"), but who is actually controlled by a gaze-contingent algorithm (see Figure 1d). This provided reciprocity in the interaction whilst enabling full experimental control. Second, we adopted a similar approach to Redcay et al. (2012) in using a cooperative social-context game in which the participant and their partner were required to fixate upon the same location – this time to catch a burglar hiding in one of six houses. Third, also similar to the approach taken by Redcay et al. (2012), we designed a baseline condition that is the same as the RJA condition except that Alan's eyes remain closed, and a green arrow guides participants to the burglar location.

Unlike RJA paradigms used in previous studies, participants in the Catch-the-Burglar paradigm are not told what role they will play at the start of each trial. Instead, they discover their social role as the trial progresses. At the start of each trial, participants see two rows of houses on a computer screen including a row of three blue doors and a row of three red doors. They are instructed to search the row of houses with blue doors (e.g., bottom row in Figure 1d) while Alan searches the row of houses with red doors (e.g., top row in Figure 1d). Participants are told that they cannot see the contents of Alan's houses and Alan cannot see the contents of their houses. Whoever finds the burglar has to guide

the other person to the correct location. The participant and Alan search their houses via fixations that open the respective doors. Thus, both partners make multiple eye-movements before the burglar is found. On RJA trials, participants discover that all of their allotted houses are empty, which indicates that the burglar must be hiding in one of Alan's houses. When Alan finds the burglar, he makes eye contact with the participant, and then guides them to the burglar by fixating on the appropriate house. Thus, participants have to wait for Alan to finish searching his houses and establish eye contact before fixating upon the relevant house. This means that participants have to monitor Alan's non-communicative gaze shifts while he completes his search, and differentiate between Alan's gaze shifts that do and do not signal joint attention bids (Cary, 1978; Senju & Johnson, 2009). To our knowledge, this is the first experimental joint attention paradigm to fully capture the intention monitoring processes involved in RJA behaviour.

In addition to establishing a more ecologically valid context in which to observe RJA behaviour, the Catch-the-Burglar paradigm includes a baseline condition that matches the RJA condition in terms of task-related demands that are not specific to RJA. This allows the elucidation of neural correlates specific to RJA using fMRI (Chapter 3; Caruana et al., 2015). On baseline trials, participants complete the same task as in the Catch-the-Burglar RJA trials except that Alan's eyes remain closed. This indicates that the "virtual interface" has been disabled. Instead, a green arrow guides participants to the burglar location in lieu of a social gaze shift. Given that our baseline condition controls for all non-social task demands involved during the RJA task, it is easier to identify neural substrates that are particularly sensitive to processes associated with RJA, and not with task differences in attention orienting or oculomotor control (cf. Saito et al., 2010; Schilbach et al., 2010; Redcay et al., 2012).

As well as measuring the neural correlates of RJA, we wanted to use the Catch-the-Burglar paradigm to measure behavioural responses in RJA trials. To our knowledge,

Redcay and colleagues (2012) have presented the only study to objectively measure RJA behaviour during an interactive experimental paradigm. They used a number of eye tracking measures to characterise behaviours across both RJA, IJA and baseline conditions in their fMRI study of joint attention (discussed above). They measured (1) the number of eye-movements made in each block of trials, (2) the duration of the joint attention experience (i.e., after the RJA or IJA behaviour had been executed), (3) the time taken for the participant to make the first eye movement at the beginning of the trial (which was not necessarily an eye movement to achieve joint attention), and (4) the time taken to achieve joint attention from the beginning of the trial (i.e., not from the point that the experimenter initiated joint attention). The fourth metric provided the most direct measure of RJA given that it measured the time taken for the RJA behaviour to be achieved.

Redcay et al. used a video interaction paradigm where the participant interacted with the experimenter in real time. Thus, the RJA response time measure could have been confounded by experimenter error or fatigue. In contrast, our new Catch-the-Burglar paradigm achieves full control over the social partner's behaviour by using virtual reality (cf. Schilbach et al. 2010). This allows the accurate measurement of saccadic reaction times (SRTs) as an index of RJA performance. Specifically, the Catch-the-Burglar paradigm measures the difference between the point in time that Alan shifts his gaze to guide the participant to the burglar, and the point in time that the participant fixates upon the burglar to achieve joint attention with Alan. This difference reflects the time taken for the participant to prepare the RJA behaviour rather than the time taken to simply execute a saccade. Thus, the Catch-the-Burglar paradigm provides an accurate and objective measure of RJA behaviour.

We have tested the reliability of the Catch-the-Burglar paradigm as a measure of RJA behaviour in two separate experiments: an fMRI study involving adults with typical development (Chapter 3; Caruana et al., 2015) and a behavioural eye tracking study that

compared performance in typically-developed adults to adults with high-functioning autism (HFA; Chapter 4). In both studies, we found an SRT effect in which adults with typical development were approximately 209 ms slower to respond to an eye gaze cue (RJA trials) than to an arrow cue (baseline trials). This suggests that, in the context of a social interaction, it takes longer to orient attention when guided by gaze cues compared to arrow cues. The consistency of this SRT effect across two independent studies suggests that the Catch-the-Burglar paradigm might be used to obtain a robust measure of RJA performance.

To ensure that the SRT effect measured by the Catch-the-Burglar task reflected differences in RJA behaviour in social and non-social conditions, and not simply differences in processing eyes versus arrows, we conducted a pilot study that compared participants' performance in two versions of the Catch-the-Burglar paradigm. Each participant completed the standard version, as outlined above, and another version that did not involve a joint search for the burglar at the start of each trial. Instead, Alan made one gaze shift to guide the participant to the correct location. This minimised the need for participants to engage in intention monitoring processes because there was only one unambiguous gaze cue on each trial (i.e., Alan only ever looked at the correct location to communicate a joint attention bid). We hypothesised that if the observed SRT effect was capturing the intention monitoring processes involved in RJA, it would be reduced in this "no-search" version of the paradigm. This is what we found. As shown in Figure 2, the discrepancy between the RJA and baseline conditions was significantly reduced when we removed the search phase, and by virtue, the intention monitoring demand. This reduction in SRTs cannot be explained by differences in the perceptual properties of gaze and arrow cues, since these stimuli were identical in both versions of the task. Thus, this pilot study demonstrated that the Catch-the-Burglar paradigm was a valid index of intention monitoring processes that pre-empt adaptive RJA behaviour (Cary, 1978).

As well as testing the validity of the Catch-the-Burglar paradigm as an index of RJA behaviour in people with typical development, we have investigated the feasibility of using the Catch-the-Burglar paradigm in people with atypical development – specifically, adults with HFA. In a study that used the Catch-the-Burglar paradigm to compare the RJA performance of adults with HFA, and control participants with typical development, we found that the SRT effect (i.e., slower responses to eye gaze than arrows) was increased in individuals with HFA compared to control participants (Chapter 4). However there were no significant group differences in performance for the baseline condition (i.e., responses to arrows). This suggests that RJA difficulties in autism may be due to a difficulty in disambiguating the communicative intent of a social partner's gaze rather than lower level impairments in non-social cognitive functions (e.g., attention orienting or oculomotor control). It also demonstrates that the Catch-the-Burglar paradigm is accessible for participants with autism.
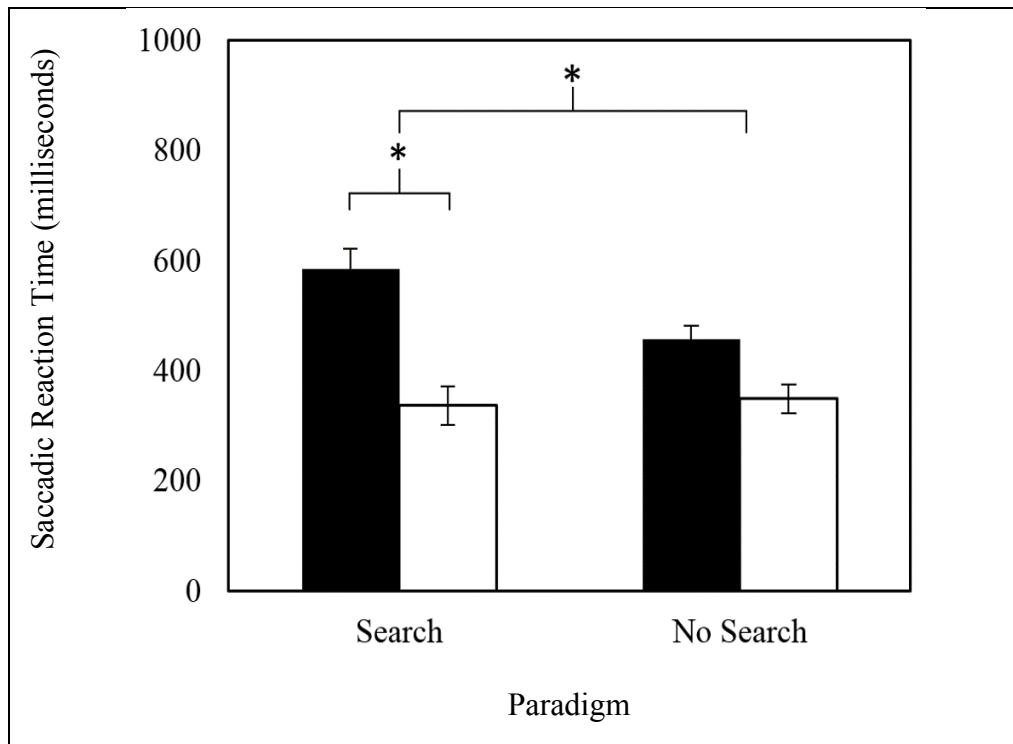
Figure 2. Saccadic reaction times for 16 adults with typical development on RJA trials

(solid bars) and baseline trials (white bars). Data is plotted separately for performance on

the original Catch-the-Burglar paradigm including the "Search" phase on each trial, and the

modified paradigm with "No Search" trials in which the virtual character only made one

gaze shift on each trial.

Initiating Joint Attention Bids

The Autism Diagnostic Observation Schedule – Generic (Lord et al., 2000)

characterises IJA behaviours as being communicative and intentional. Specifically, IJA is

defined as protodeclarative (to show or share) rather than protoimperative (to request;

Bruinsma, Koegel, & Koegel, 2004). In typical development, infants begin to initiate joint

attention (IJA) between six and twelve months of age, usually after the emergence of RJA

behaviours (Mundy et al., 1994). In autism, RJA impairments often recede during

development (Mundy et al., 1994), but IJA impairments often persist well into adolescence

(Hobson & Hobson, 2007).

Compared to RJA, relatively few studies have investigated IJA using experimental

paradigms. Some studies using fMRI or electroencephalography (EEG) have employed

tasks that require participants to saccade to a particular location and then evaluate whether their partner has followed their gaze (e.g., Lachat, Hugueville, Lemarechal, Conty, & George, 2012; Schilbach et al., 2010). Whilst these studies have certainly asked participants to engage in IJA behaviour in IJA trials, they have also asked participants to engage in IJA behaviour in corresponding "baseline" trials. Subtracting brain responses in the latter from the former reveals neural responses associated with evaluating the success of IJA behaviour (i.e., EAJA) rather than IJA *per se*. Thus, these studies will be discussed in the following section that focuses on EAJA.

In contrast to these EAJA studies, the aforementioned study by Redcay et al. (2012) required participants to engage in IJA behaviours in IJA trials but not in baseline trials. Participants were instructed to search for a mouse whose tail was protruding from behind one of four cheeses displayed on the screen. Joint attention was initiated by fixating upon the correct location, whereupon their partner, "Lee", fixated upon the same location. Brain activity triggered in IJA trials was compared to baseline trials in which participants were instructed to locate a mouse while Lee's eyes were closed. A strength of this paradigm was that subtracting activation in baseline trials from IJA trials identified IJA-related brain activation that controlled for activation associated with performing the visual search and making eye movements. However, participants knew before each trial whether they would be initiating or responding to a joint attention bid. Thus, on IJA trials, participants knew that Lee would follow them, and there was no requirement for them to make eye contact with her before initiating joint attention. Thus, participants could have completed the task whilst ignoring Lee. This differs from IJA behaviours in real social interactions that often depend upon a person making eye contact with their partner to ascertain that their partner is paying attention before a bid for joint attention is attempted.

To our knowledge, the Catch-the-Burglar paradigm is the first to capture this attention monitoring process in IJA (Chapter 3; Caruana et al., 2015). As described above,

participants are not instructed about their social role before each trial. Rather, they are required to monitor Alan' gaze behaviour to identify opportunities for joint attention. Like RJA trials, at the beginning of each IJA trial, both the participant and Alan search separate rows of houses looking for a burglar (the search phase). If the participant finds the burglar, they then have to establish eye contact with Alan to ensure that they have his attention before guiding him to the burglar location (i.e., initiate joint attention). Thus, participants are required to monitor Alan's attention and convey their intent to communicate by waiting for him to return eye contact before initiating joint attention.

In order to isolate the neural correlates specific to IJA in applications of the Catch-the-Burglar paradigm in fMRI studies, we have developed a non-social baseline condition for IJA trials. In this condition, participants fixate upon a small grey circle superimposed on Alan's nasion (analogous to fixating on his eyes). When it turns green (analogous to waiting for Alan to make eye contact), the participant is required to saccade back to the correct location (as they would do if they were guiding Alan). Thus, participants have to make the same sequence of eye-movements in both IJA and baseline trials. Subtracting activation in the baseline trials from IJA trials effectively controls for activation related to non-social task demands present in IJA trials, thus allowing the neural correlates specific to IJA to be measured.

In addition to measuring the neural correlates of IJA, the Catch-the-Burglar paradigm has been used to measure IJA behaviour. It provides two eye-tracking measures of IJA performance. First, it measures target dwell time, which is the period of time that passes between participants fixating upon the burglar and the time that they attempt to establish eye contact with Alan. This provides a metric of the participant's readiness to obtain their partner's attention so that they can initiate a bid for joint attention. Second, the paradigm measures the frequency with which participants attempt to initiate joint attention before they establish eye contact with Alan. Performance on these measures during IJA

trials can be contrasted with performance during trials in the closely matched baseline condition. To our knowledge, this is the first joint attention paradigm to allow ecologically-valid eye-tracking measures of IJA accuracy and efficiency.

Evaluating the Achievement of Joint Attention (EAJA)

A third component of joint attention is the ability to evaluate the achievement of joint attention (EAJA). This is a particularly important process since an initiator of joint attention must determine whether they have been successful in guiding their partner's attention. Emery (2000) argues that when two people are mutually aware that they have achieved joint attention, this becomes a separate social phenomenon called "shared attention". However, we consider EAJA to be a component of joint attention, partly because EAJA only requires awareness from one person who initiates the bid for joint attention, and partly because EAJA is considered to be an aspect of "joint attention" within the existing experimental literature (cf. Pfeiffer et al., 2013).
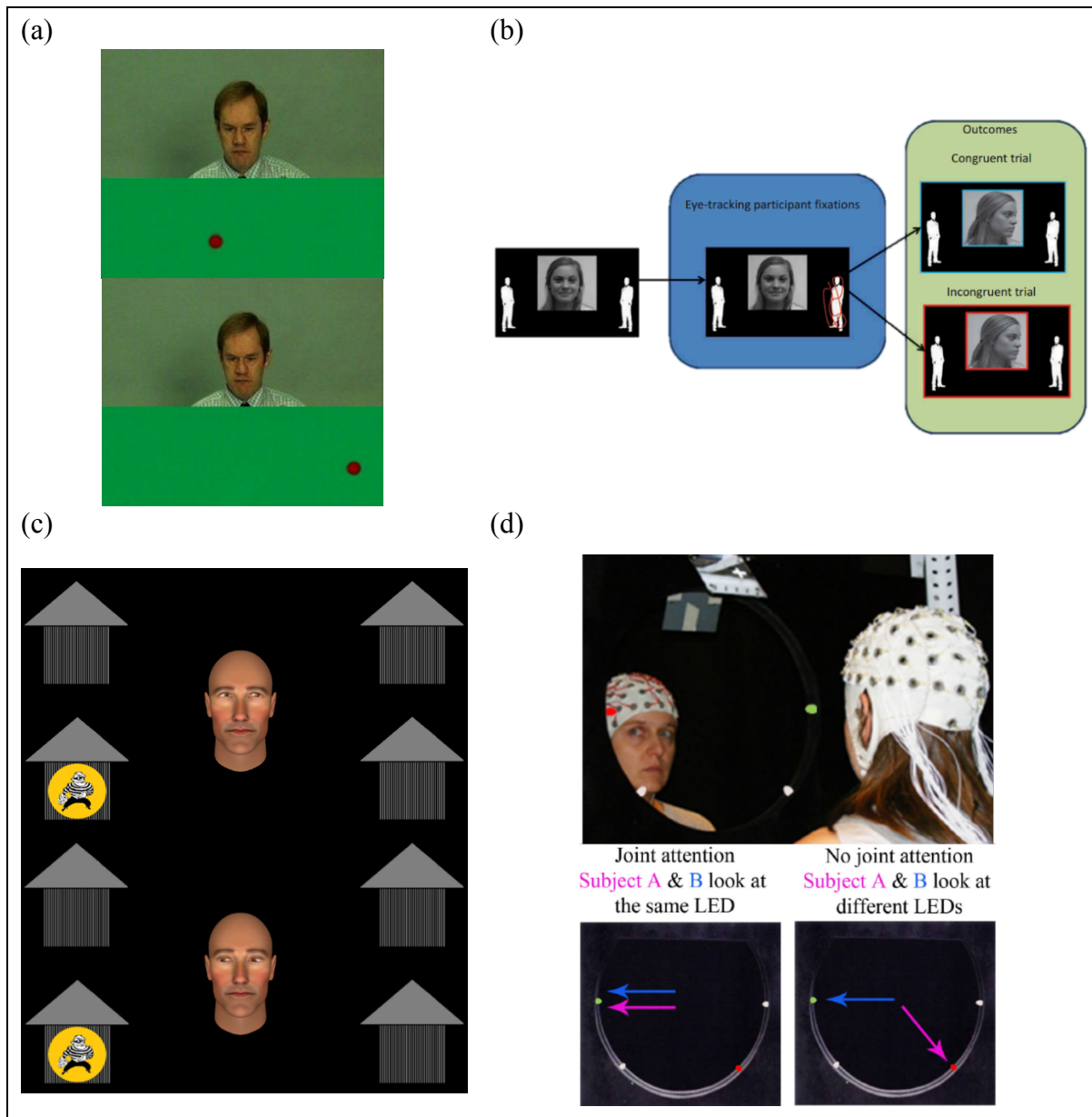
Figure 3. Still frames from interactive paradigms measuring EAJA. Each panel depicts trials where joint attention was achieved, and not achieved. (a) Williams et al.'s (2005) video paradigm. (b) Gordon et al.'s (2013) gaze-contingent video paradigm. (c) Caruana, de Lissa, & McArthur's (2015) cooperative virtual reality paradigm (d) Lachat et al.'s (2012) EEG hyperscanning paradigm. All figures reprinted with permissions.

A number of studies have investigated the neural correlates associated with EAJA using gaze congruency tasks. These tasks have typically asked participants to fixate upon an object under two conditions: (1) a "joint attention" condition in which a participant's social partner follows their gaze to achieve joint attention (congruent trials); and (2) a "no joint attention" condition in which the participant's social partner does not followed the participant's gaze, and looks instead at another location to avoid joint attention (incongruent trials).

Studies using fMRI and EEG measures have compared neural responses under these two conditions to elucidate where and when EAJA is processed in the brain. For example, in an fMRI study by Williams, Waiter, Perra, Perrett, and Whiten (2005), participants were presented with a video recording of a red dot presented to either the bottom-left or bottom-right of the screen as well as a man at the top of the screen (see Figure 3a). Participants were required to look at the red dot. Increased activation in the ventral mPFC was found in congruent trials when the man in the video also looked at the dot (joint attention) compared to incongruent trials in which he looked in the opposite direction (no joint attention). This suggested that EAJA may be supported by brain regions that have been previously associated with representing the mental perspectives of others (Amodio & Frith, 2006).

Similarly, Gordon, Eilbott, Feldman, Pelphrey, and Vander Wyk (2013) presented participants with a display that comprised a central video frame that was flanked by two human silhouettes to the left and right (see Figure 3b). The video frame depicted the upper torso and face of a female named "Sally". Participants were asked to fixate upon one of the silhouettes. Once a fixation was detected, a video of Sally was played to depict her turning her head congruently to look at the same location (joint attention) or incongruently (no joint attention). Observing the achievement of joint attention resulted in greater activation of the ACC, right fusiform gyrus, right amygdala, striatum and parahippocampus,

compared to trials where Sally responded incongruently to avoid joint attention.

The EAJA-related brain regions identified by Gordon et al. (2013) were different to those identified by Williams et al. (2005). This might have occurred because Gordon et al. specifically instructed participants to "guide" Sally to look at one of the silhouettes, whilst Williams et al. simply instructed participants to look at the red dot. Furthermore, the different outcomes of these studies might have occurred because Gordon et al. used a gaze-contingent video display to ensure that the video of Sally only started when the participant fixated on one of the silhouettes, while Williams et al. presented video clips that played independently to participants' gaze behaviour. Thus, Gordon et al.'s paradigm may have been more ecologically valid than Williams et al.'s paradigm by presenting participants with reciprocal social behaviours resulting in the activation of different brain regions. Nevertheless, even Gordon et al.'s study was limited in ecological validity because participants were aware that the videos were pre-recorded and hence that they were not interacting with a real person. According to the second person approach, an ecologically valid EAJA paradigm should ensure that participants believe that they are genuinely engaged in a social interaction.

In contrast to Gordon et al.'s (2013) paradigm, the paradigm developed by Schilbach et al. (2010; see above) ensured that participants believed that they were engaging in a genuine social interaction – even though their virtual partner was controlled by a gaze-contingent algorithm (see Figure 1c). Schilbach et al.'s paradigm comprised four types of trials: (1) on OTHER_JA trials participants responded congruently to their partner's joint attention bid (i.e., RJA condition described earlier), (2) on OTHER_NOJA trials participants responded incongruently to their partner's joint attention bid (i.e., RJA baseline condition described earlier), (3) on SELF_JA trials participants fixated on one of three squares located above and beside the virtual character's face, and their partner responded congruently to achieve joint attention, and (4) on SELF_NOJA trials

participants fixated upon one of three squares and their partner responded incongruently by saccading to a different square. Schilbach and colleagues found that SELF_JA was associated with increased activation of the ventral striatum compared to SELF_NOJA trials. Schilbach et al. interpreted the activation of the ventral striatum as reflecting the hedonic experience of evaluating the successful outcome of IJA (i.e., EAJA). This is consistent with other studies that have observed activation in this region during reward processing paradigms (McClure, York, & Montague, 2004).

In contrast to Schilbach et al.'s findings, Williams et al. (2005) did not identify activation in the ventral striatum associated with EAJA. A key difference between these studies was the degree to which participants believed that they were interacting with a real human being. This suggests that brain activation in EAJA trials may be affected by a participant's belief about whether they are interacting with a real human being. We have evidence to support this interpretation from a series of event-related potential (ERP) studies that employed a modified version of the Catch-the-Burglar paradigm (Chapter 5 & 6; Caruana, de Lissa, & McArthur, 2015). In this modified paradigm, participants interacted with Alan, whose face was depicted in the centre of a screen, along with four cartoon buildings in each corner of the screen (see Figure 3c). Each building represented an exit from a prison complex. Participants were told that Alan was a "prison guard" who was in charge of patrolling the inside of the prison. It was his job to "lock down" any exits that were breached by escaping prisoners. Participants were told that they were a "watch person" who was in charge of patrolling the outside of the prison. They were told that a convicted burglar would try to escape on each trial, and that it was their job to fixate upon the appropriate exit so that Alan could lock it down. After fixating upon the relevant exit, participants were required to fixate back upon Alan's face to determine if he had achieved joint attention. To ensure that participants observed Alan's response, Alan only responded once a participant had fixated back on his face.

Participants were informed that Alan might sometimes be distracted by scuffles within the prison. Accordingly, on 50% of trials, Alan responded congruently to the participant's joint attention bid, and on 50% of trials he responded incongruently by gazing towards an incorrect location (see Figure 3c). The difference between ERP responses to congruent and incongruent trials revealed the neural correlates of EAJA.

Using this modified version of the Catch-the-Burglar paradigm, we found that centro-parietal P350 peaks were larger and later for Alan's incongruent gaze shifts than Alan's congruent (joint attention) gaze shifts (Chapter 5; Caruana, de Lissa, et al., 2015). This P350 effect was absent in a second sample of participants who completed the same task but were told that Alan was controlled by a computer algorithm (Chapter 6). This suggests that beliefs about the authenticity of a social interaction influences the neural processing of EAJA. Thus, joint attention paradigms should simulate interactions that participants believe to be genuine (cf. Schilbach et al. 2010).

An outstanding question that is not addressed by the aforementioned studies is whether the neural measures of EAJA are influenced by whether joint attention is achieved incidentally (cf. Williams et al., 2005) or intentionally (cf. Chapter 5; Caruana, de Lissa, et al., 2015). Lachat et al. (2012) addressed this question by examining the effect of incidental and intentional joint attention experiences on the neural processing of EAJA. They used EEG to record the neural activity of two people simultaneously as they interacted with each other face-to-face. The high temporal resolution of EEG allowed the measurement of neural processes at the point in time that joint attention had been achieved. Participants interacted with each other face-to-face through a circular viewing window that had LED lights positioned around its circumference (see Figure 3d). On "social instruction" trials, one participant was required to look at an LED light of their choice. The other participant was instructed to either follow their partner's gaze to achieve joint attention, or to intentionally avoid joint attention by fixating the LED light on the opposite

side of the viewing window. On "colour instruction" trials, both participants were told to look at an LED light of a particular colour. This resulted in the incidental achievement of joint attention on some trials but not on others. Analyses of EEG data focused on alpha-mu suppression, which has been associated with empathising (Perry, Bentin, Bartal, Lamm, & Decety, 2010), mentalising (Pineda & Hecht, 2009), and social coordination (Naeem, Prasad, Watson, & Kelso, 2012). Achieving joint attention was associated with a power decrease in parieto-occipital alpha and centro-parietal mu oscillations (i.e., 11-13 Hz) regardless of colour instruction or social instruction. Whilst the reliability of this null effect needs to be confirmed in replication studies, it suggests that the neural processing of EAJA may not be sensitive to whether joint attention is achieved incidentally or intentionally (Lachat et al., 2012). Further work is needed to clarify whether participants must intentionally pursue joint attention in order to validly measure the neural processes of EAJA.

Recommendations for the Measurement of Joint Attention

A second person approach to measuring joint attention demands both ecological validity (via monitoring of a social partner's communicative intentions, simulating reciprocity between the participant and their social partner, and supporting the experience of genuine social interactions in real time) and experimental control (via affording full control over the visual and temporal dynamics of the social stimuli, and employing adequate baseline conditions). However, as demonstrated by the review above, there appears to be some degree of tension between these two requirements. We will now discuss developments and recommendations for future second person approaches that attempt to measure joint attention within truly interactive contexts (cf. Schilbach et al., 2013). Specifically, we discuss the need for paradigms that simulate social interactions that are realistically complex, motivate intentional communicative behaviours, and are genuinely engaging. These methodological issues are of value not only to the investigation

of joint attention, but to the application of second person approaches to social cognition and neuroscience research in general.

Complex Interactions

One of the aims of the second person approach is to measure social cognition while people participate in dynamic interactions that simulate the complexity of everyday experiences (Schilbach et al., 2013). For instance, real interactions involve multiple modalities of communication (e.g., gaze, expression, gesture, posture, and speech) in environments that can be dynamic, distracting, and unpredictable (Conty, Dezecache, Hugueville, & Grèzes, 2012). However, experimental measures that have attempted to distil joint attention behaviours within simulated interactions do not necessarily capture this complexity. This may explain why individuals with autism have been found to have difficulties in joint attention and social communication in naturalistic settings (Mundy, Sullivan, & Mastergeorge, 2009), but not necessarily in experiments involving simple and repetitive social judgements or behaviours (Nation & Penny, 2008).

Most previous experimental studies of joint attention have employed tasks in which the joint attention "event" is isolated and unambiguous. For instance, RJA trials typically involve a single gaze cue to which participants are required to respond (Redcay et al., 2012; Schilbach et al., 2010; Saito et al., 2010, Lachat et al., 2012). Likewise, IJA trials ask participants to make a single eye-movement towards a particular location with full knowledge that their partner will follow (e.g., Redcay et al., 2012). These tasks allow the maintenance of experimental control and a large number of relatively brief trials. However, they do not capture a number of cognitive processes that would typically be used to navigate complex social interactions, such as decisions about (1) which gaze shifts are communicative, (2) whether to respond to, or initiate, a joint attention bid, (3) whether our social partner is willing to communicate, or (4) whether our social partner is intentionally communicating with us.

As outlined above, the search phase of our interactive Catch-the-Burglar paradigm addresses some of these issues. Rather than overtly instructing participants about whether to engage in RJA or IJA on each trial (cf. Lachat et al., 2012; Redcay et al., 2012, Saito et al., 2010; Schilbach et al., 2010), participants intuitively determine their role depending on whether they find the burglar (IJA trials) or not (RJA trials). In RJA trials, the participant has to differentiate between Alan's averted gaze shifts that are intended to guide them to the burglar and those that are being made as Alan completes his search. Likewise, in IJA trials, the participant cannot assume that their partner will follow every gaze shift they make. Rather, participants must establish eye contact with Alan before engaging in IJA. Whilst previous paradigms have overtly instructed participants to establish eye contact (e.g., Schilbach et al., 2010), the Catch-the-Burglar paradigm is the first to ensure that eye contact is essential to the achievement of joint attention. Using eye contact in this way is an important aspect of intentional IJA in everyday interactions as it informs the initiator that they have their partner's attention, and signals the initiator's intent to communicate (Cary, 1978; Senju & Johnson, 2009).

Whilst we have taken several steps towards the development of a joint attention paradigm that closely simulates the complexity of everyday social interactions, more can be done. For instance, all experimental studies of joint attention to date have measured joint attention behaviour within nonverbal gaze-based interactions. This is a sensible starting point for a number of conceptual and practical reasons. However, in real social interactions, joint attention can also be achieved using other communicative signals (e.g., pointing). Future studies could investigate how social cues from multiple modalities are simultaneously used during joint attention episodes. However, it is noteworthy that the maintenance of experimental control will become a greater challenge as tasks become increasingly complex.

Intentional Interactions

The second person approach aims to develop paradigms in which social cognition is engaged in the same way as it would be in an everyday social interaction. Typically, we engage in joint attention behaviours to pursue a current goal (e.g., request assistance, share information, communicate interest, or understand the visual perspective of our social partner). Thus, to understand the cognitive and neural mechanisms that support these behaviours, it could be argued that joint attention should be studied within an intentional context (Tomasello, 1995). This is especially relevant to studies of autism, given the emphasis on intentional join attention behaviours in theoretical accounts of the disorder (Adamson et al., 2009; Dawson et al., 2004) and clinical assessments (Lord et al., 2000).

In most of the experimental joint attention paradigms discussed above, the participant was not engaged in an intentional interaction with their partner. In some tasks, joint attention was achieved incidentally while the participant was following another cue that happened to lead to them to look in the same location as their partner (e.g., Williams et al., 2005). In other tasks, participants were under instruction to guide or follow their partner without an intuitive social reason to do this beyond the explicit instruction (e.g., Gordon et al., 2014; Lachat et al., 2012; Saito et al., 2010; Schilbach et al., 2010). In contrast, Redcay et al. (2012) asked participants to intentionally engage in joint attention to communicate information that was relevant to their current goal (i.e., catching a mouse). We have adopted a similar approach with the Catch-the-Burglar paradigm, which naturally motivates participants to engage in intentional RJA and IJA behaviours in order to locate and capture a burglar.

Engaging Interactions

Another important aim of the second person approach is to enable participants to experience social interactions that are genuinely engaging. Some joint attention studies have not met this requirement because participants have responded to, or are directed to, a

computer-programmed animation or video recording of a character whom they know is not interacting with them in real time (e.g., Gordon et al., 2014; Williams, et al., 2005). Other studies of joint attention have met this requirement since they have involved genuine social interactions in which individuals interact with another person either face-to-face (Lachat et al 2012) or via a live video feed (Redcay et al., 2012; Saito et al. 2010). However, this has been at the expense of experimental control over various aspects of the interaction (e.g., timing of behaviour, display of ostensive facial expressions, and control over stimulus attractiveness). Virtual reality offers an innovative tool for overcoming this challenge (Wilms et al., 2010). Virtual characters provide a social partner whose behaviour and appearance can be subject to full experimental control (Georgescu, Kuzmanovic, Roth, Bente, & Vogeley, 2014) whilst allowing participants to experience the "copresence" that characterises genuine social interactions (Biocca, Harms, & Burgoon, 2003). Pioneering virtual reality studies have attempted to achieve the experience of copresence by deceiving participants to believe that their virtual partner is controlled by the online eye-movements of another human being (cf. Schilbach et al., 2010; Wilms et al., 2010).

An important question surrounding the use of virtual reality is whether it is necessary for participants to believe that their social partner is real or not. Our aforementioned ERP studies suggest that this belief does matter (Chapters 5 & 6; Caruana, de Lissa, et al., 2015). As noted earlier, participants who believed that they interacted with a real person showed a significantly larger P350 response when their virtual partner followed their joint attention bid compared to when their partner failed to follow their joint attention bid. This P350 effect was not found in individuals who believed that the virtual character was controlled by a computer program. Notably, we also found that beliefs about the virtual character's agency had a similar effect on the earlier N170 responses to gaze shifts made by the virtual character. The N170 response has been associated with the perceptual processing of gaze shifts (see Itier & Batty, 2009 for a review). These findings

are consistent with earlier studies employing social cognition tasks which have revealed that neural processing is significantly modulated by participants' beliefs about whether they are interacting with a human agent rather than a computer (Gallagher, Jack, Roepstorff, & Frith, 2002; McCabe, Houser, Ryan, Smith, & Trouard, 2001). They also align with the Intentional Stance Model, which proposes that mentalising mechanisms (i.e., the cognitive process of understanding the mental states of others) may only be recruited when individuals believe that they are interacting with a real person, and that these mentalising mechanisms have a top-down effect on the processing of social information, such as gaze shifts (Wykowska, Wiese, Prosser, & Müller, 2014).

If we accept the importance of convincing participants that they are engaged in a genuine social interaction, we should also consider how this belief can be established in the laboratory. Clearly, the virtual partner should display behaviours that are realistically human (Georgescu et al., 2014). For example, temporal jitter may be added to the gaze behaviour displayed by the virtual character to make it feel less robotic (Wilms et al., 2010). As well as including temporal jitter, the Catch-the-Burglar paradigm ensures that Alan' gaze behaviour varies across trials in terms of the number of eye movements that he makes during his search, and the order in which he searches his allotted houses. The paradigm also ensures that Alan performs the task in a manner that is consistent with him being controlled by another human being. For example, if a participant guides Alan to the wrong location, he is programmed to follow them (e.g., Chapter 3; Caruana et al., 2015). In addition, Alan is fallible. In our ERP studies investigating EAJA, Alan was programmed to make a large number of errors (so we could contrast achieving versus failing to achieve joint attention). We used a cover story to explain Alan's errors by telling participants that Alan was occasionally distracted by scuffles within the prison (e.g., Chapter 5; Caruana, de Lissa, et al., 2015). This cover story ensured that Alan's performance did not appear unrealistically poor or uncooperative.

Previous studies that have used virtual reality to simulate joint attention experiences have also recognised the importance of using cover stories to convince participants that the virtual character is being controlled by a real human in real-time (e.g., Schilbach et al., 2010; Wilms et al., 2010). However, the nature of the cover story has varied. Some studies have used relatively elaborate cover stories with confederates to convince participants that a virtual character is controlled by a human (Schilbach et al., 2010; Wilms et al., 2010). However, in our studies applying the Catch-the-Burglar paradigm, the subjective ratings provided by participants after the experiment have indicated that participants were deceived that Alan represented a real person without the need of a human confederate. Rather, we simply told participants that they were interacting with another person, and provided an explanation of how the virtual interface supposedly worked. Further work is required to determine which features of the task are necessary for deceiving participants into believing that they are interacting with another human being. This will inform the most ethical, effective, and practical induction of these beliefs.

Conclusions

The aim of this chapter was to merge the strengths of previous second person joint attention paradigms within a new experimental paradigm that provided an ecologically valid measure of the cognitive and neural mechanisms of RJA, IJA, and EAJA. We have also discussed a number of challenges associated with designing joint attention paradigms that provide full experimental control as well as ecological validity. Finally, we have suggested a number of techniques to overcome these challenges. To date, virtual reality has offered the most promising methodology in which to overcome these challenges, but it is still in its infancy. With increased development in coming years, virtual reality should provide new insights into the cognitive and neural processes that support our ability to understand and communicate with others. This is of great importance to the fields of social cognition and social neuroscience, and to the empirical investigation of the social-

communication difficulties that characterise autism spectrum disorders.

References

Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. A. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of Autism & Developmental Disorders, 39*(1), 84-96. doi: 10.1007/s10803-008-0601-7

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268-277. doi: 10.1038/nrn1884

Baldwin, D. A. (2014). Understanding the link between joint attention and language. In C. Moore & P. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development*: Taylor & Francis.

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

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: review and suggested criteria. *Presence: Teleoperators and Virtual Environments, 12*(5), 456-480. doi: 10.1162/105474603322761270

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation & Developmental Disabilities Research Reviews, 10*(3), 169-175. doi: 10.1002/mrdd.20036

Bruner, J. S. (1974). From communication to language, a psychological perspective. *Cognition, 3*(3), 255-287. doi: 10.1016/0010-0277(74)90012-2

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

Caruana, N., de Lissa, P., & McArthur, G. (2015). The neural time course of evaluating self-initiated joint attention bids. *Brain and Cognition*. doi: http://dx.doi.org/10.1016/j.bandc.2015.06.001

Cary, M. S. (1978). The Role of Gaze in the Initiation of Conversation. *Social Psychology, 41*(3), 269-271. doi: 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. doi: 10.1098/rstb.2002.1199

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., & Drew, A. (1997). Infants with autism: An investigation of empathy, pretend play, joint attention, and imitation. *Developmental Psychology, 33*(5), 781-789. doi: 10.1037/0012-1649.33.5.781

Chawarska, K., Klin, A., & Volkmar, F. (2003). Automatic attention cueing through eye movement in 2-year-old children with autism. *Child Development, 74*(4), 1108-1122. doi. 10.1111/1467-8624.00595

Conty, L., Dezecache, G., Hugueville, L., & Grèzes, J. (2012). Early binding of gaze, gesture, and emotion: neural time course and correlates. *The Journal of Neuroscience, 32*(13), 4531-4539. doi: 10.1523/jneurosci.5636-11.2012

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. *Developmental Psychology, 40*(2), 271-283. doi: http://dx.doi.org/10.1037/0012-1649.40.2.271

Dawson, G., Webb, S., Schellenberg, G. D., Dager, S., Friedman, S., Aylward, E., & Richards, T. (2002). Defining the broader phenotype of autism: Genetic, brain, and behavioral perspectives. *Development and Psychopathology, 14*(3), 581-611. doi: 10.1017/S0954579402003103

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*(6), 581-604. doi:10.1016/S0149-7634(00)00025-7

Friesen, C., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490-495. doi: 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. doi: 10.1037/0033-2909.133.4.694

Frith, U. (2003). *Autism: Explaining the enigma* (2nd ed.). Malden, MA: Blackwell.

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage, 16*(3), 814-821. doi:10.1006/nimg.2002.1117

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., & Vogeley, K. (2014). The use of virtual characters to assess and train nonverbal communication in high-functioning autism. *Frontiers in Human Neuroscience, 8*. doi: http://dx.doi.org/10.3389/fnhum.2014.00807

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. doi: 10.1080/17470919.2013.832374

Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: a longitudinal study of gaze following during interactions with mothers and strangers. *Developmental Science, 13*(6), 839-848. doi: 10.1111/j.1467-7687.2009.00945.x

Gredebäck, G., Theuring, C., Hauf, P., & Kenward, B. (2008). The microstructure of infants' gaze as they view adult shifts in overt attention. *Infancy, 13*(5), 533-543. doi: 10.1080/15250000802329529

Happe, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience, 9*(10), 1218-1220.

Hobson, J., & Hobson, P. (2007). Identification: The missing link between joint attention and imitation? *Development and Psychopathology, 19*(02), 411-431. doi: doi:10.1017/S0954579407070204

Hobson, R. P. (2008). Interpersonally situated cognition. *International Journal of Philosophical Studies, 16*(3), 377-397. doi: 10.1080/09672550802113300

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*(2), 131-134. doi: 10.1111/1467-9280.00024

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. doi: http://dx.doi.org/10.1016/j.neubiorev.2009.02.004

Johnson, M. H., Griffin, R., Csibra, G., Halit, H., Farroni, T., De Haan, M., Tucker, L. A., Baron-Cohen, S., & Richards, J. (2005). The emergence of the social brain network: Evidence from typical and atypical development. *Development and Psychopathology, 17*(03), 599-619. doi: doi:10.1017/S0954579405050297

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*(13), 1897-1916. doi: 10.1016/0042-6989(94)00279-U

Lachat, F., Hugueville, L., Lemarechal, J. D., Conty, L., & George, N. (2012). Oscillatory brain correlates of live joint attention: a dual-EEG study. *Frontiers in Human Neuroscience, 6*, 156. doi: 10.3389/fnhum.2012.00156

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction
detection: A dissociation between geometric and joint attention skills in autism.
British Journal of Developmental Psychology, 15(1), 77-95. doi: 10.1111/j.2044-
835X.1997.tb00726.x

Leekam, S. R., Lopez, B., & Moore, C. (2000). Attention and joint attention in preschool
children with autism. *Developmental Psychology, 36*, 261-273. doi:
http://dx.doi.org/10.1037/0012-1649.36.2.261

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. *Journal of Autism and Developmental Disorders, 30*(3), 205–223. doi:
10.1023/A:1005592401947

Loveland, K. A., & Landry, S. H. (1986). Joint attention and language in autism and
developmental language delay. *Journal of Autism & Developmental Disorders,
16*(3), 335-349. doi: 10.1007/BF01531663

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging
study of cooperation in two-person reciprocal exchange. *Proceedings of the
National Academy of Sciences, 98*(20), 11832-11835. doi: 10.1073/pnas.211415698

McClure, S. M., York, M. K., & Montague, P. R. (2004). The neural substrates of reward
processing in humans: the modern role of fMRI. *The Neuroscientist, 10*(3), 260-
268. doi: 10.1177/1073858404263526

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A. E., & Seibert, J. M. (2003). *A
manual for the Abridged Early Social Communication Scales (ESCS)*. Coral
Gables, FL, University of Miami.

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

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20*(1), 115-128. doi: 10.1007/bf02206861

Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. *Development and Psychopathology, 6*(03), 389-401. doi: doi:10.1017/S0954579400006003

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research, 2*(1), 2-21. doi: 10.1002/aur.61

Murray, D. S., Creaghead, N. A., Manning-Courtney, P., Shear, P. K., Bean, J., & Prendeville, J.-A. (2008). The relationship between joint attention and language in children with autism spectrum disorders. *Focus on Autism & Other Developmental Disabilities, 23*(1), 5-14. doi: 10.1177/1088357607311443

Naeem, M., Prasad, G., Watson, D. R., & Kelso, J. A. (2012). Electrophysiological signatures of intentional social coordination in the 10-12 Hz range. *NeuroImage, 59*(2), 1795-1803. doi: 10.1016/j.neuroimage.2011.08.010

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? Is it automatic? Is it social? *Development and Psychopathology, 20*(01), 79-97. doi: doi:10.1017/S0954579408000047

Navab, A., Gillespie-Lynch, K., Johnson, S. P., Sigman, M., & Hutman, T. (2011). Eye-tracking as a measure of responsiveness to joint attention in infants at risk for autism. *Infancy*, 416-431. doi: 10.1111/j.1532-7078.2011.00082.x

Okada, T., Sato, W., Murai, T., Kubota, Y., & Toichi, M. (2003). Eye gaze triggers visuospatial attentional shift in individuals with autism. *Psychologia, 46*(4), 246-254. doi: 10.2117/psysoc.2003.246

Osterling, J., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology, 14*(02), 239-251. doi: doi:10.1017/S0954579402002031

Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia, 41*(2), 156-170. doi: 10.1016/s0028-3932(02)00146-x

Perry, A., Bentin, S., Bartal, I. B., Lamm, C., & Decety, J. (2010). "Feeling" the pain of those who are different from us: Modulation of EEG in the mu/alpha range. *Cognitive Affective and Behavioral Neuroscience, 10*(4), 493-504. doi: 10.3758/cabn.10.4.493

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience and Biobehavioural Reviews, 37*(10), 2516-2528. doi:10.1016/j.neubiorev.2013.07.017

Pineda, J. A., & Hecht, E. (2009). Mirroring and mu rhythm involvement in social cognition: are there dissociable subcomponents of theory of mind? *Biological Psychology, 80*(3), 306-314. doi: 10.1016/j.biopsycho.2008.11.003

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

Redcay, E., Dodell-Feder, D., Mavros, P. L., Kleiner, A. M., Pearrow, M. J., Triantafyllou, C., Gabrieli, J. D., & Saxe, R. (2012). brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. *Human Brain Mapping, 34,* 2511-2523. doi: 10.1002/hbm.22086.

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research, 24*(3), 715-718. doi: 10.1016/j.cogbrainres.2005.02.007

Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., Uchiyama, H., Kosaka, H., Okazawa, H., Fujibayashi, Y., & Sadato, N. (2010). "Stay tuned": inter-individual neural synchronization during mutual gaze and joint attention. *Frontiers in Integrative Neuroscience, 4*, 127. doi: 10.3389/fnint.2010.00127

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393-414. doi: 10.1017/S0140525X12000660.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah N. J., Fink, G. R., & Vogeley, K. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience, 22*(12), 2702-2715. doi: 10.1162/jocn.2009.21401.

Seibert, J. M., Hogan, A. E., & Mundy, P. C. (1982). Assessing interactional competencies: The early social-communication scales. *Infant Mental Health Journal, 3*(4), 244-258. doi: 10.1002/1097-0355(198224)3:4<244::aid-imhj2280030406>3.0.co;2-r

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology, 18*(9), 668-671. doi: 10.1016/j.cub.2008.03.059

Senju, A., & Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences, 13*(3), 127-134. doi: http://dx.doi.org/10.1016/j.tics.2008.11.009

Stone, W. L., Ousley, O. Y., & Littleford, C. D. (1997). Motor imitation in young children with autism: what's the object? *Journal of Abnormal Child Psychology, 25*, 475-485. doi: 10.1023/A:1022685731726

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development*. Hillsdale: Lawrence Erlbaum Associates

Tylén, K., Allen, M., Hunter, B. K., & Roepstorff, A. (2012). Interaction vs. observation: distinctive modes of social cognition in human brain and behavior? A combined fMRI and eye-tracking study. *Frontiers in Human Neuroscience, 6*, 331. doi: 10.3389/fnhum.2012.00331

Vlamings, P. H. J. M., Stauder, J. E. A., van Son, I. A. M., Mottron, & Laurent. (2005). Atypical visual orienting to gaze- and arrow-cues in adults with high functioning autism. *Journal of Autism and Developmental Disorders, 35*(3), 267-277. doi: 10.1007/s10803-005-3289-y

Volkmar, F. R., Cohen, D. J., Bregman, J. D., Hooks, M. Y., & Stevenson, J. M. (1989). An examination of social typologies in autism. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*(1), 82-86. doi:10.1097/00004583-198901000-00015

von Hofsten, C., Dahlström, E., & Fredriksson, Y. (2005). 12-month-old infants' perception of attention direction in static video images. *Infancy, 8*(3), 217-231. doi: 10.1207/s15327078in0803_2

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. doi: 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. doi: 10.1093/scan/nsq024

Wong, C., & Kasari, C. (2012). Play and joint attention of children with autism in the preschool special education classroom. *Journal of Autism and Developmental Disorders, 42*(10), 2152-2161. doi: 10.1007/s10803-012-1467-2

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. doi: 10.1371/journal.pone.0094339

Chapter 3

A Frontotemporoparietal Network Common to Initiating and Responding to Joint Attention Bids

Nathan **Caruana**[123], Jon **Brock**[124] & Alexandra **Woolgar**[123]

[1] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[2] ARC Centre of Excellence in Cognition and its Disorders, Sydney

[3] Perception in Action Research Centre, Sydney

[4] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

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.

A Frontotemporoparietal Network Common to

Initiating and Responding to Joint Attention Bids

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, Sigman, & Kasari, 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, Koegel, & Koegel, 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 & Adamson, 1984; Scaife & Bruner, 1975), while IJA develops later, at approximately 12 months of age (Bates, Benigni, Bretherton, Csmaioni, & Volterra, 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, Sigman, & Kasari, 1994). Contrastingly, IJA impairments often persist well into adolescence and adulthood (Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998; Lord et al., 2000; MacDonald et al., 2006; Mundy & Jarrold, 2010; Mundy, Sigman, Ungerer, & Sherman, 1986; Sigman & Ruskin, 1999).

According to the Parallel and Distributed-Processing model (PDPM) of joint attention (Mundy & Jarrold, 2010; Mundy & Newell, 2007; Mundy, Sullivan, & Mastergeorge, 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 goal-directed attention towards stimuli which signal rewarding experiences (anterior cingulate cortex; Mundy & Newell, 2007). This frontal network is also thought to include the medial prefrontal cortex (mPFC; Amodio & Frith, 2006; Williams, Waiter, Perra, Perrett, & Whiten, 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 & 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 representations 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 in which participants interacted with an avatar who they believed was depicting the eye movements of a social partner outside the scanner (see Wilms et al., 2010). 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 & 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 & Raine, 2003; Simpson, Snyder, Gusnard, & Raichle, 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, Derrfuss, & Von Cramon, 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 (i.e., 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 (Figure 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 (i.e., 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 & Newell, 2007; Redcay et al., 2012; Redcay et al., 2010; 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

79

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, Hlushchuk, Hari, & Schürmann, 2009; Ramsey, Hansen, Apperly, & Samson, 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 coloured irises as eyes with light irises tend to be difficult to calibrate in scanner environments (Gordon, Eilbott, Feldman, Pelphrey, & Vander Wyk, 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 (Figure 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 & 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.

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

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 Figure 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 Materials 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 Figure 1), while the avatar was always responsible for searching the houses with red doors (e.g., the top row in Figure 1). The blue doors appeared in the top or bottom row of houses, counterbalanced within participants and across acquisition runs. This prevented confounds driven by saccade trajectory, since downward saccades have previously been found to differentially recruit frontal regions (Tzelepi, Laskaris, Amditis, & Kapoula, 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 (Figure 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 initiated bid for 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 distribution between 500-1000 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 (Figure 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 (Figure 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 0-2 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, Figure 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; Figure 3, second column) and initiating (IJAc; Figure 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 "Supplementary video 1" (online, doi:10.1016/j.neuroimage.2014.12.041).
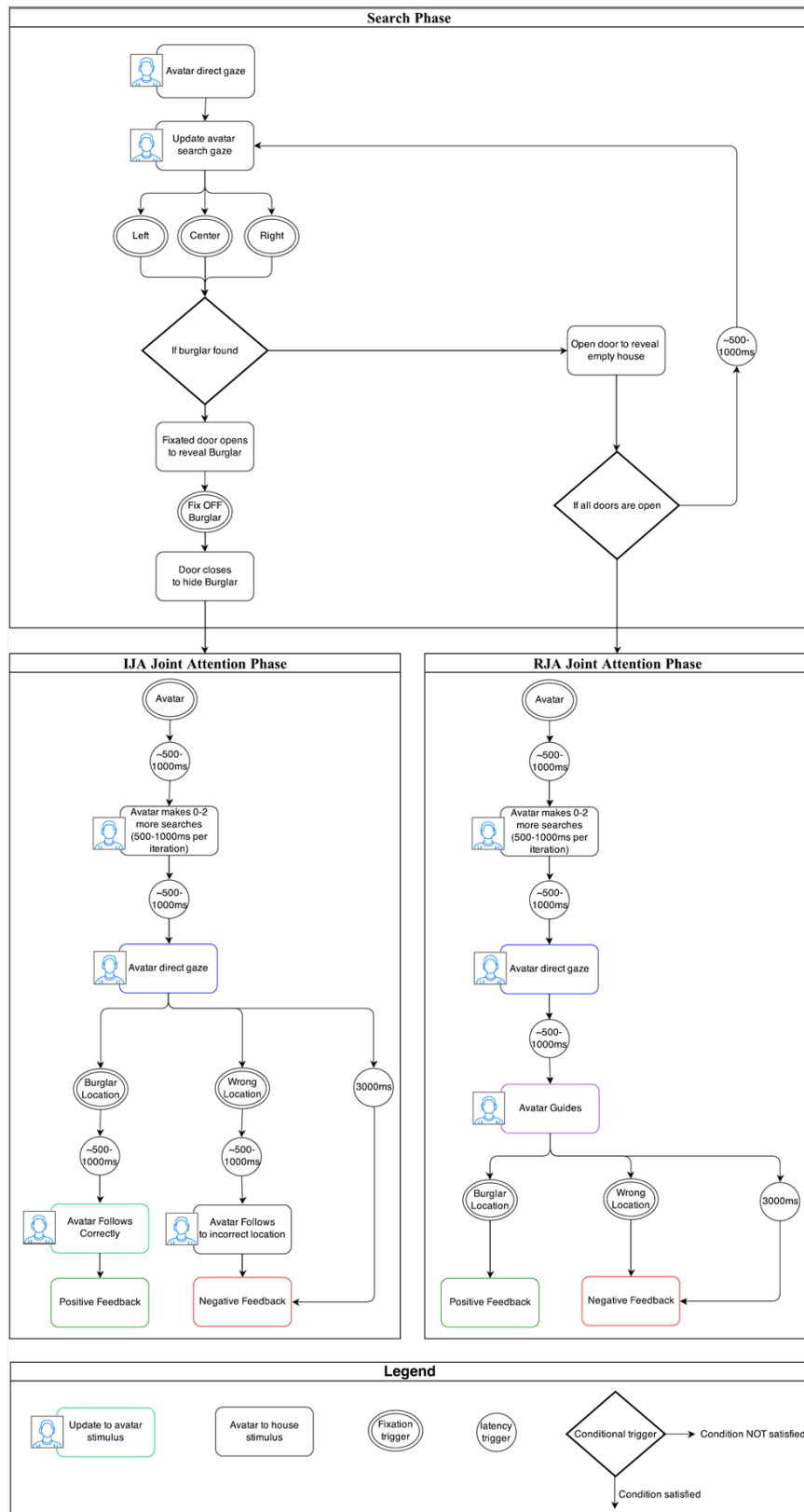
Figure 2. Schematic summary of interactive algorithm. This algorithm was the same for the social and non-social conditions, apart from the central stimulus that is used (e.g., avatar direct gaze versus green fixation point). This diagram has been labelled to reflect the stimuli in the social conditions.
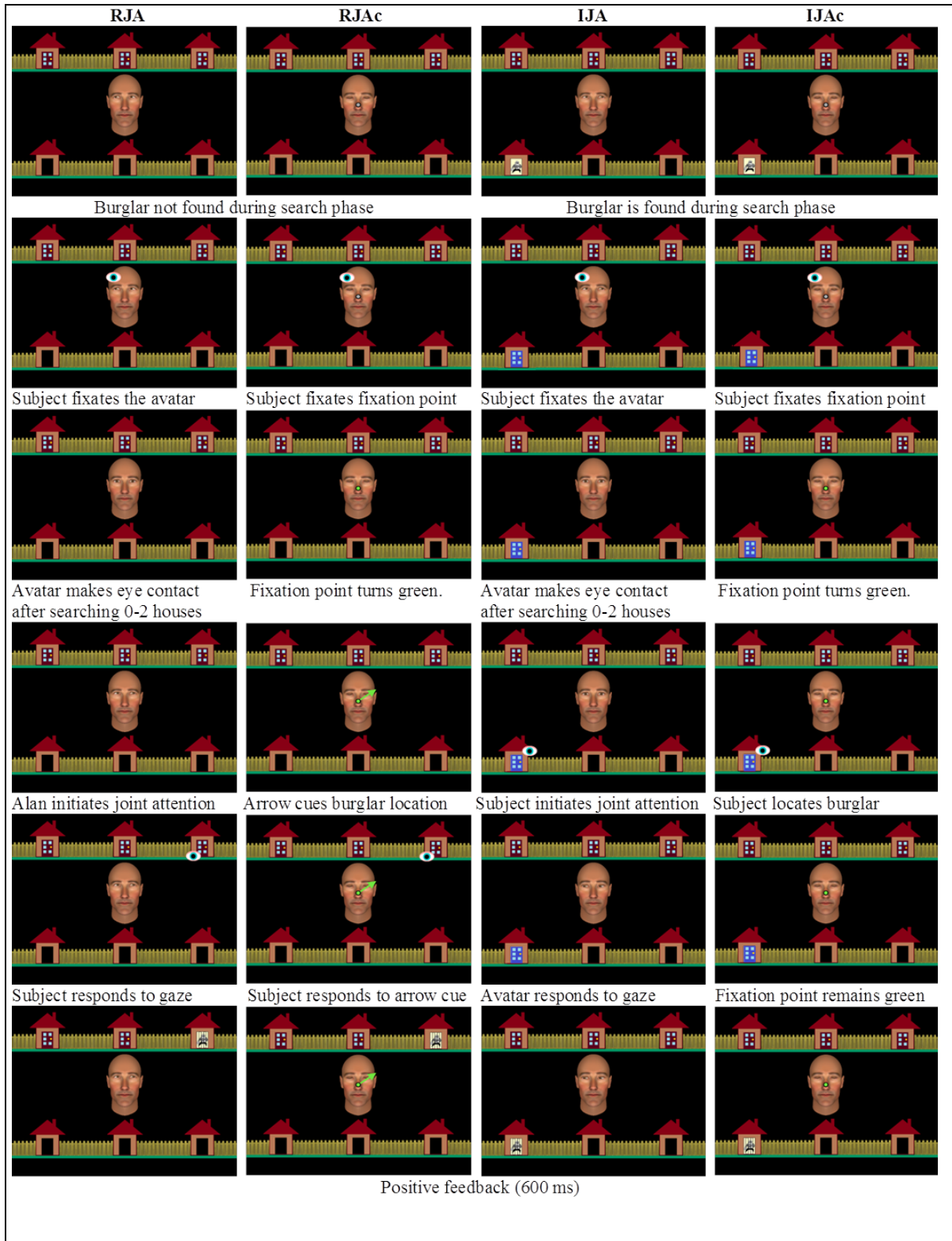
Figure 3. Schematic representation of trial sequence by condition. Eye symbol represents the location of the

participant's gaze and was not visible to the participant.

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 with text 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 comprising 108 trials. Each run consisted of 27 trials of each condition (i.e., RJA, RJAc, IJA, 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 co-operative they thought Alan was. Participants were debriefed about the true nature of the social interaction. They were told that they were not interacting with a person named 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 1000Hz 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 (Figure 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 3T 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 = 64^2; voxel size = 3.0 x 3.0 x 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 – 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 x 1.0 x 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 image, 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 Figure 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, Wager, Lindquist, Ferrera, & Hirsch, 2008; Henson, 2006; Vogeley et al., 2004; Woolgar, Golland, & Bode, 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 unbalanced 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 RJA 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 a 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).

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).

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 measured ANOVA revealed a significant interaction effect of social context (social versus control) and social role (responding versus 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 times 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 and intentions 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 GAOIs on each trial (Figure 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 Figure 4a-d.

Table 2.
Neural Correlates

Region	(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
Frontal																												
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
	L	46	-26	44	14	44	4.36	R	46	34	54	16	57	4.77														
								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	46	8	44	48	2387	8.49								L	6	-28	2	60	301	5.12
																						R	6	6	36	50	12	3.96
Insula	L	47	-34	22	2	96	5.26	R	47	26	22	-6	455	5.05														
Inferior Frontal Gyrus	R	44	48	22	18	3075	5.55	R	45	42	24	8	455	6.73	R	44	54	22	26	16	4.13	R	45	48	36	8	10	4.63
Superior Frontal Gyrus								L	6	-22	6	60	2387	8.98								B	6	20	6	58	12	4.03
*																						R	6	18	12	68	58	8.22
SMA																						B	6	4	8	58	42	4.39
Precentral Gyrus								L	6	-30	-2	58	2387	8.07								L	6	-36	-4	50	301	5.65
Anterior Cingulate Cortex								B	32	4	40	12	27	4.84														
Middle Cingulate Cortex								R		10	18	32	27	4.45								R		12	-32	46	11	4.06
								L		-6	-6	32	32	7.53								L	23	-4	-8	32	24	5.69
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							
Temporal																												
Inferior Temporal Gyrus	R	20	52	-22	-14	52	5.04	R	37	40	-52	-14	27	4.50								R	37	56	-58	-4	1471	6.61
Middle Temporal Gyrus								R	22	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	40	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																					
Temporal Pole																						L	38	-30	6	-26	14	4.43
Fusiform Gyrus								R	V4	38	-60	-12	27	4.45														
Parietal																												
Posterior Cingulate Cortex								B	29	8	-40	14	36	5.61														
Precuneus	B	7	4	-62	58	355	5.18	B		10	-62	51	5137	8.15	B	5	2	-52	54	93	4.06	B		10	-54	46	80	5.48
Occipital																												
Cuneus								B	18	4	-84	24	5137	8.57														
Calcarine Gyrus								B	17	0	-96	-2	13	4.41								B	18	-4	-80	14	6118	11.66
Lingual Gyrus								L	19	-22	-66	-4	64	6.58														
Superior Occipital Gyrus								R		18	-82	32	5137	7.36														
Middle Occipital Gyrus								L	V5	-46	-74	0	5137	7.60								R		30	-74	30	6118	9.59
																						L	V5	-46	-74	0	6118	8.83
Subcortical																												
Thalamus								R		12	-12	6	187	9.88								R	18	12	-12	6	21	5.13
Amygdala*	B	34	16	-6	-10	21	4.67																					
Cerebellum								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. 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)]. Extent threshold=10. *Cluster labels based on nearest grey matter to which the significant cluster extended. H= Hemisphere , BA=Brodmann Area.
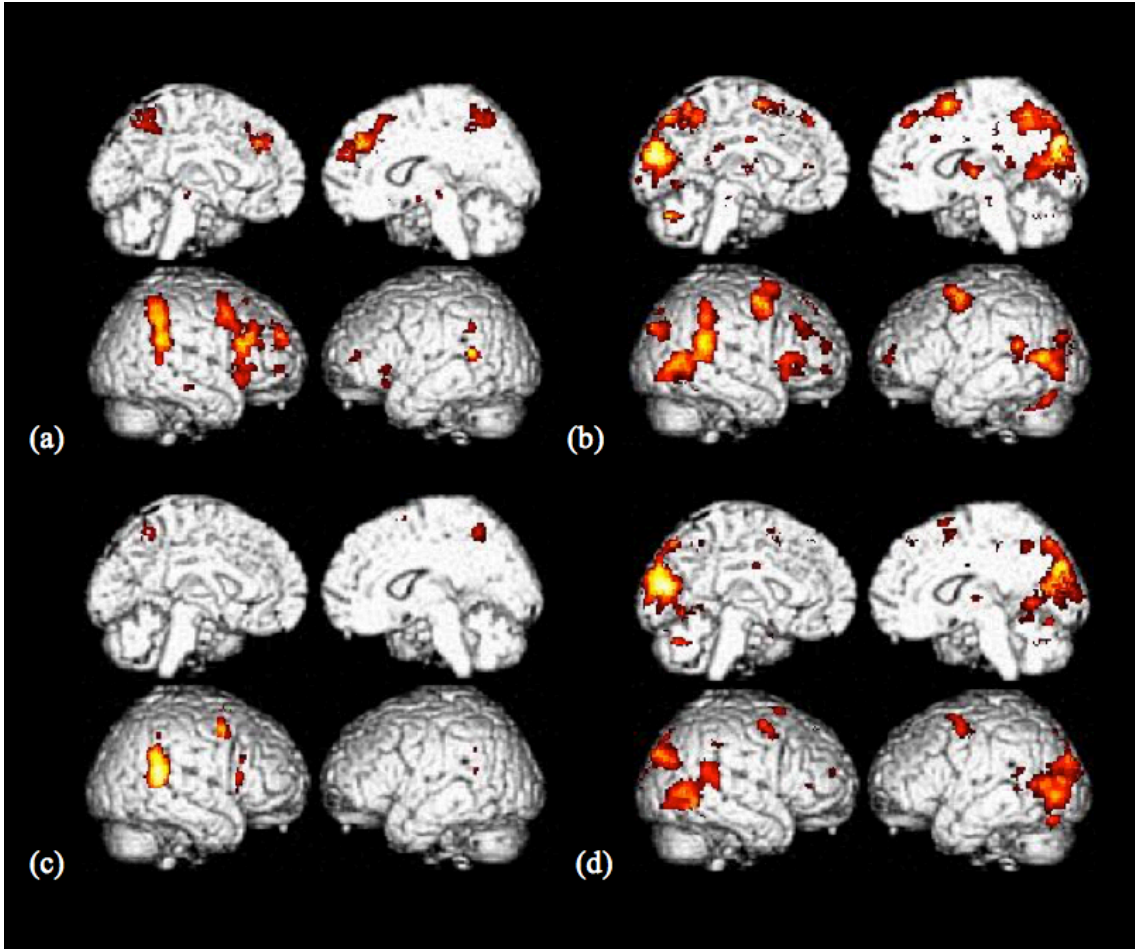
Figure 4. Thresholded statistical parametric maps showing activity associated with (a) Responding to joint attention (*RJA-RJAc*) (b) Initiating joint attention (*IJA – IJAc*) (c) Activation common to Responding and Initiating (d) Initiating over and above Responding [*(IJA - IJAc) - (RJA - RJAc)*]. 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-IJAc)*].

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, Figure 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, Figure 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, Figure 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, Figure 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 & Cunningham, 2009). Also see Supplementary Materials 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 & 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 & 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 themself, and that of the person that they are interacting with (Mundy et al., 2009, p. 7). 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, Apperly, Chiavarino, & Humphreys, 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, Abrams, Astafiev, Shulman, & Corbetta, 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 and colleagues (2010) reported increased activation in rTPJ during trials when joint attention was not achieved (i.e., NO_JA 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, Schulz, & Jiang, 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, Pelphrey, & McCarthy, 2008; Pelphrey, Morris, & McCarthy, 2004; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 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, Kimmig, Glauche, & Haller, 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.

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 –

IJAc), IJA resulted in significantly more activation in a bilateral frontotemporoparietal

network than RJA; (IJA – IJAc) – (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 & Newell, 2007; Picard & 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 – IJAc 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 (IJAc) 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 and intentions 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, Patel, & Shulman, 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 & 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 & 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, Wager, Taylor, & Liberzon, 2002) and self-agency; that is perceiving an outcome as resulting from one's own actions, versus that of another (Farrer & 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.

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268-277. doi: 10.1038/nrn1884

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interactions. *Child Development, 55*(4), 1278. doi: 10.1111/1467-8624.ep7302943

Bates, E., Benigni, L., Bretherton, I., Csmaioni, L., & Volterra, V. (1979). The emergence of symbols: cognition and communication in infancy. New York: Academic Press.

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. doi:10.1016/j.neuropsychologia.2004.06.018

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research, 10*(3), 169-175. doi: 10.1002/mrdd.20036

Bruner, J. S. (1974). From communication to language, a psychological perspective. *Cognition, 3*(3), 255-287. doi: 10.1016/0010-0277(74)90012-2

Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal–parietal junction. *Trends in Cognitive Sciences*, *17*(7), 328-336. doi: http://dx.doi.org/10.1016/j.tics.2013.05.007

Cary, M. S. (1978). The role of gaze in the initiation of conversation. *Social Psychology - Germany*, *41*(3), 269-271. doi: 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. doi:10.1016/j.neuron.2008.04.017

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. *Developmental Psychology, 40*(2), 271-283. doi.org/10.1037/0012-1649.40.2.271

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. doi: 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. doi: http://dx.doi.org/10.1006/nimg.2001.1009

Gallese, V. (2001). The 'shared manifold' hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies, 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. *Social Neuroscience - UK*, 1-11. doi: 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. doi: 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. doi: http://dx.doi.org/10.1016/j.neuroimage.2009.01.063

Henson, R. N. A. (2006). Efficient Experimental Design for fMRI. In K. J. Friston, A. T. Ashburner, S. J. Kiebel, T. E. Nichols & W. D. Penny (Eds.), Statistical parametric mapping: The analysis of functional brain images: The analysis of functional brain images. London: Elsevier.

Henson, R. N. A. (Producer). (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 B. B. Lahey & T. E. Moffitt (Eds.), Causes of conduct disorder and juvenile delinquency (pp. 277-304). New York: Guilford.

Kimball, S., & Mattis, P. (1995). GIMP Image Manipulation Program (Version 2.8.2). Berkley.

Kincade, 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. *Journal of Neuroscience, 25*(18), 4593-4604. doi: 10.1523/jneurosci.0236-05.2005

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience, 4*(4), 423-428. doi: 10.1093/scan/nsp052

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. *Journal of Autism and Developmental Disorders, 30*(3), 205–223. doi: 10.1023/A:1005592401947

Marchetti, I., & Koster, E. H. W. (2014). Brain and intersubjectivity: A Hegelian hypothesis on the self-other neurodynamics. [Opinion]. *Frontiers in Human Neuroscience, 8*. doi: 10.3389/fnhum.2014.00011

Mitchell, J.P., 2008. Activity in right temporo−parietal junction is not selective for theory−of−mind. *Cerebral Cortex, 18*(2), 262–271. doi: 10.1093/cercor/bhm051

Morris, J. P., Pelphrey, K. A., & McCarthy, G. (2008). Perceived causality influences brain activity evoked by biological motion. *Social Neuroscience - UK, 3*(1), 16-25. doi: 10.1080/17470910701476686

Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks, 23*(8), 985-997. doi: 10.1016/j.neunet.2010.08.009

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

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20*(1), 115-128. doi: 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 Research, 2*(1), 2-21. doi: 10.1002/aur.61

Ohlendorf, S., Kimmig, H., Glauche, V., & Haller, S. (2007). Gaze pursuit, 'attention pursuit' and their effects on cortical activations. *European Journal of Neuroscience, 26*(7), 2096-2108. doi: 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. *Journal of Cognitive Neuroscience, 16*(10), 1706-1716. doi: 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. doi: 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. *Cerebral Cortex*, *13*(9), 977-

986. doi: 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. *Journal of Cognitive Neuroscience, 25*(5), 670-684.

doi:10.1162/jocn_a_00345

Redcay, E., Dodell-Feder, D., Mavros, P. L., Kleiner, A. M., Pearrow, M. J., Triantafyllou,

C., Gabrieli, J. D., & Saxe, R. (2012). Atypical brain activation patterns during a

face-to-face joint attention game in adults with autism spectrum disorder. *Human

Brain Mapping, 34,* 2511-2523. doi: 10.1002/hbm.22086.

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. doi:

10.1016/j.neuroimage.2010.01.052

Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind

reasoning: Implications for autism. *Brain Cognition, 55*(1), 209-219. doi:

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. *Nature

Neuroscience, 7*(5), 499-500. doi:10.1038/nn1223

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. doi: 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. *Social Neuroscience - UK, 1*(3-4), 284-298. doi: 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. doi: 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. *Nature, 253*(5489), 265-266. doi:10.1038/253265a0

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. *Social Cognition and Affective Neuroscience*, *6*(4), 393-403. doi: 10.1093/scan/nsq067

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36,* 441- 462 doi: 10.1017/S0140525X12002452

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah N. J., Fink, G. R., & Vogeley, K. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience, 22*(12), 2702-2715. doi: 10.1162/jocn.2009.21401.

Senju, A., & Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Science*, *13*(3), 127-134. doi: 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. *Nature Reviews Neuroscience, 12*(3), 154-167. doi: 10.1038/nrn2994

Sigman, M., Ruskin, E., 1999. Continuity and change in the social competence of children with autism, Down syndrome and developmental delays. *Monographs of the Society for Research in Child Development, 64* (1), 1–114. doi: 10.1111/1540-5834.00002

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 perfromance. *Proceedings of the National Academy of Sciences USA, 98*, 683-687. doi: 10.1073/pnas.98.2.683

Singular Inversions. (2008). FaceGen Modeller (Version 3.3) [Computer Software]. Toronto, ON: Singular Inversions.

SR Research. (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 Research, 1321*, 105-116. doi: 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. *Journal of Cognitive Neuroscience, 16*(5), 817-827. doi: 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. doi: 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 Cognition and Affective Neuroscience, 5*(1), 98-107. doi: 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*. doi: http://dx.doi.org/10.1016/j.neuroimage.2014.04.059

Chapter 4

Joint Attention in Adults with High-Functioning Autism

Nathan **Caruana**[123], Heidi Stieglitz **Ham**[4], Jon **Brock**[125,] Alexandra **Woolgar**[123],

Nadine **Kloth**[26], Romina **Palermo**[26], & Genevieve **McArthur**[12]

[1] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[2] ARC Centre of Excellence in Cognition and its Disorders

[3] Perception in Action Research Centre, Sydney

[4] School of Occupational Therapy and Social Work, CURTIN UNIVERSITY, Perth

[5] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

[6] School of Psychology, UNIVERSITY OF WESTERN AUSTRALIA, Perth

Abstract

Joint attention is a cognitive ability that supports social communication and the regulation of interpersonal relationships. Individuals with autism demonstrate impairments in both initiating and responding to joint attention bids in naturalistic settings (e.g., Hobson & Hobson, 2007; Mundy, Sigman, & Kasari, 1994). However, this phenomenon has not yet been tested in an experimental setting using an ecologically valid and experimentally controlled paradigm. In the current study, we used such a paradigm to test joint attention in 17 adults with high functioning autism (HFA) and 17 age-matched controls with typical development. We also tested these participants on a non-social paradigm that had similar task demands. Adults with HFA were initially slower to respond to joint attention bids than controls. However, by the end of testing, they achieved a level of performance commensurate to controls. This learning effect was largely driven by "systemisers" in the HFA group, who had a stronger tendency to understand complex phenomena as rule-governed systems. The responses of adults with HFA to our new ecologically-valid and experimentally-controlled paradigm are discussed in relation to the development of new social cognition interventions.

Joint Attention in Adults with High-functioning Autism

Introduction

Joint attention is the ability to achieve a common focus of attention with another person during a social interaction, and is crucial to the development of language and social learning (Adamson, Bakeman, Deckner, & Romski, 2009; Baron-Cohen, 1995; Charman, 2003; Mundy, Sigman, & Kasari, 1990; Murray et al., 2008; Tomasello, 1995). In a joint attention episode, person A initiates a joint attention bid by intentionally guiding person B's attention towards an object or event. Joint attention is achieved if person B responds by following person A's communicative eye movements (Bruinsma, Koegel, & Koegel, 2004). In typical development, responding and initating behaviours begin to emerge at approximately six months (Bakeman & Adamson, 1984; Scaife & Bruner, 1975) and twelve months respectively (Bates, Benigni, Bretherton, Csmaioni, & Volterra, 1979). However, in children with autism, responding behaviours do not begin to emerge until cognitive development is equivalent to that of 30-36 months of typical development (Mundy et al., 1994), and impairments in initiating joint attention often persist well into adolescence (e.g., Charman, 2003; Hobson & Hobson, 2007; Mundy, Sigman, Ungerer, & Sherman, 1986; Sigman & Ruskin, 1999). Individual variation in joint attention is also a reliable predictor of social functioning (Lord et al., 2000; Stone, Ousley, & Littleford, 1997) and language development in children on the autism spectrum (Charman, 2003; Dawson et al., 2004; Mundy et al., 1990).

Joint attention impairments in autism have typically been identfied in observational studies of natural and semi-structured social interactions (Charman et al., 1997; Dawson et al., 2004; Loveland & Landry, 1986; Mundy et al., 1990; Osterling & Dawson, 1994; Osterling, Dawson, & Munson, 2002; Wong & Kasari, 2012). Experimental studies of joint attention in autism typically use gaze-cueing paradigms (see Frischen, Bayliss, & Tipper, 2007, for a review). These studies often employ variations of the Posner-cueing

121

paradigm (Posner, 1980), in which a target presented on the left or right side of the screen is preceded by a central eye gaze cue (e.g., a pair of eyes) directing attention either to the target location or in the opposite direction. Gaze-cueing studies have provided inconsistent evidence for joint attention impairments in autism (see Nation & Penny, 2008 for review). Whilst some have reported reduced or absent cue validity effects in autism (i.e., faster detection of targets preceded by valid cues rather than invalid cues; e.g., Johnson et al., 2005; Ristic et al., 2005), others have not (e.g., Chawarska, Klin, & Volkmar, 2003; Kylliainen & Hietanen, 2004; Okada, Sato, Murai, Kubota, & Toichi, 2003; Vlamings, Stauder, van Son, Mottron, & Laurent, 2005).

There are three possible explanations for these inconsistent findings: (1) subtle differences in stimuli and experimental procedure (see Frischen et al., 2007, for reviews; Nation & Penny, 2008); (2) heterogeneity within the autism population (Brock, 2011); and (3) insensitivity of gaze-cueing paradigms due to low ecological validity. Regarding (3), gaze-cueing paradigms measure the influence that task-irrelevant gaze stimuli have on the participant's ability to detect a target on each trial. That is, they measure the extent to which individuals "reflexively" respond to gaze cues. However, in real-life interactions, it is not appropriate to follow every gaze shift made by a social partner. Instead, one needs to orient to gaze shifts that are likely to be meaningful or communicative (Cary, 1978). Accordingly, successful joint attention behaviours rely on the ability to use eye contact as a cue to determine whether gaze shifts are meaningful, and if a social partner is ready to communicate (Cary, 1978). It is possible that joint attention impairments in autism may be associated with difficulty in understanding the meaning of eye contact rather than an inability to orient to gaze (Senju & Johnson, 2009). This is consistent with evidence that individuals with autism are less effective in using eye contact to understand the goals and actions of others (Phillips, Baron-Cohen, & Rutter, 1992) or to assess the relevance of an upcoming gaze shift (Böckler, Timmermans, Sebanz, Vogeley, & Schilbach, 2014).

The aim of the current study was to use a new experimentally controlled and ecologically valid eye-tracking paradigm to assess the behaviours of responding to joint attention bids (RJA) and initiating joint attention bids (IJA) in people with autism. Adults with high-functioning autism (HFA) and adults with typical development (i.e., controls) were asked to interact with an animated virtual character (avatar) that they believed was being controlled by another person. Together they played a cooperative game in which they had to catch a burglar that was hiding inside one of six houses displayed on the screen. Each trial began with a search phase in which the participant and avatar both searched their allotted houses. Whoever found the burglar had to guide their partner to that location by first making eye contact and then gazing at the location of the burglar. This created a social condition that (1) elicited intentional RJA and IJA behaviours, (2) informed participants of their social role (i.e., responder or initiator) throughout the course of each trial without overt instruction, and (3) required participants to use eye contact as a cue to identify joint attention opportunities. Performance on RJA and IJA trials was compared with performance in control conditions that had the same task demands but did not involve social interaction (RJAc and IJAc).

Using a number of performance metrics, we investigated whether participants with HFA performed the responding and initiating tasks as well as control participants, and whether any group differences were due to social context (i.e., poor performance in the RJA or IJA conditions only) or due to task demands (i.e., poor performance in both RJA and RJAc or IJA and IJAc conditions). If we found that adults with HFA were impaired in responding or initiating behaviours, we also aimed to determine (1) if adults with HFA with the poorest joint attention performance engaged in more "systemising" (i.e., the propensity to understand complex phenomena as rule-governed systems; Baron-Cohen, 2008); and (2) if adults with HFA with high systemising tendencies made the greatest improvements in joint attention task performance across trials. If (1) proved true, we

predicted that (2) would also be true given that high-systemisers are more likely to implement strategies that facilitate performance with practice on the task (Wheelwright et al., 2006).

Method

Ethical Statement

The study was approved by the Human Research Ethics Committee at Macquarie University (MQ; reference number: 5201200021) and ratified by the University of Western Australia (UWA). Participants received payment or course credit for their time and provided written consent before participating.

Participants

Eighteen adults with HFA were tested at UWA (Perth, Australia). Each had a formal diagnosis of Autism Spectrum Disorder confirmed by a clinical psychologist. The Ritvo Autism Asperger Diagnostic Scale-Revised (RAADS-R; Ritvo et al., 2011) was administered. All participants with HFA scored above the threshold of 65 except one participant who scored 60. Nonverbal IQ was assessed using the Matrices subtest of the Kauffman Brief Intelligence Test – Second Edition (KBIT-2; Kaufman & Kaufman, 2004). One participant was excluded because their nonverbal IQ score was below 85. This resulted in a final sample of 17 adults with HFA (6 female, M_{age} = 26.43, SD=14.53, $M_{Nonverbal\ IQ}$ = 105.94, SD = 13.45, $M_{RAADS-R}$ = 126.44, SD = 25.47). Their performance was compared to 17 control participants with typical development who were tested at MQ (Sydney, Australia; 6 female, M_{age} = 26.47, SD=11.86, $M_{Nonverbal\ IQ}$ = 105.70, SD = 12.46, $M_{RAADS-R}$ = 53.06, SD = 25.06). No control participant scored above threshold on the RAADS-R. The two groups comprised the same numbers of females and males, and did not differ significantly in age or nonverbal IQ (see Table 1).

Participants completed a series of questionnaires, including the Autism Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), the Empathising

Quotient (EQ), and the Systemising Quotient (SQ; Wheelright et al., 2006). One

participant with HFA declined to complete these questionnaires. Participant demographic

and questionnaire data for each group is shown in Table 1. Adults in the HFA group scored

significantly higher on the AQ, and lower on the EQ, compared to controls. No significant

group differences were found on the SQ.

Table 1.

Demographic and Questionnaire Data

	HFA group		Control group		Statistics	
	M	*SD*	*M*	*SD*		
Age	26.47	11.86	26.43	14.53	$t(32) = 0.01$	$p = .993$
Nonverbal IQ	105.94	13.45	105.70	12.46	$t(32) = 0.05$	$p = .958$
RAADS-R	126.44	25.47	53.06	25.06	$t(31) = 9.87$	$p < .001$
AQ	27.81	11.08	9.24	6.31	$t(31) = 5.96$	$p < .001$
EQ	27.31	10.93	55.35	10.69	$t(31) = -7.45$	$p < .001$
SQ	66.44	25.47	53.05	25.06	$t(31) = 1.52$	$p = .139$

Note. Nonverbal IQ scores were based on the standard score obtained using the KBIT-2 Matrices subtest

(Kaufman & Kaufman, 2004). Total raw scores are reported for the Ritvo Autism Asperger Diagnostic Scale-

Revised (RAADS-R; Ritvo et al., 2011), Autism Quotient (AQ; Baron-Cohen, et al., 2001), Empathising

Quotient (EQ), and Systemising Quotient (SQ; Wheelright et al., 2006).

Stimuli

We employed an interactive paradigm previously used to isolate the neural

correlates of responding to, and initiating joint attention (Caruana, Brock, & Woolgar,

2015). The stimuli comprised an anthropomorphic avatar face presented in the centre of

the screen that subtended 6.5 degrees of visual angle. The avatar's eyes were manipulated

using *FaceGen* (Singular Inversions, 2008) so that his gaze could be directed either at the

participant or towards one of the six houses that were presented on the screen (see Figure 1). The houses, which each subtended four degrees of visual angle, were arranged in two horizontal rows above and below the avatar.
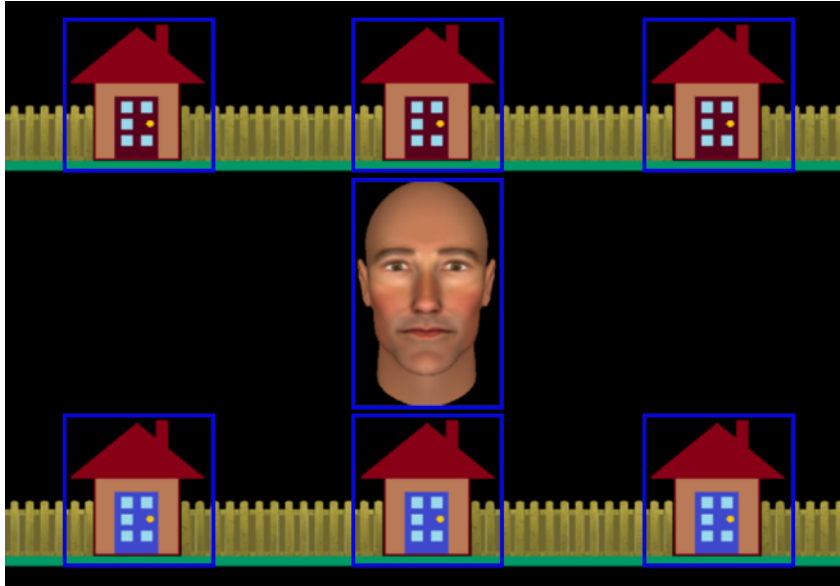
Figure 1. Gaze areas of interest (GAOIs), represented by blue rectangles, overlaid on participants' view of stimuli.

Joint Attention Task

Social conditions (RJA and IJA). In the social conditions, participants played a cooperative game with an avatar whom they believed represented the gaze behaviour of another person named "Alan" who was in a nearby eye tracking laboratory. Participants were told that Alan would control the avatar's gaze using live-infrared eye tracking over a high-speed network. In reality, a gaze-contingent algorithm used the online recordings of the participant's eye movements to program the avatar's responsive behaviour (see Caruana et al., 2015, for a description of this algorithm). The joint goal of the participant and Alan was to catch a burglar that was hiding inside one of the six houses that were presented on the screen.

RJA

RJAc

Subject does not find burglar during search

Subject fixates on avatar to make eye contact (or on grey fixation point)

Avatar completes search (looks at 0-2 more houses) and makes eye (or fixation point turns green)

Avatar's gaze (or arrow) indicates the burglar's location

ANALYSIS PERIOD (A)

Saccadic Reaction Time. Latency between cue and onset of responsive saccades

Subject must make a saccade to the burglar location

Feedback. Burglar appears behind bars at the burglar's location

IJA

IJAc

Subject finds burglar during search

ANALYSIS PERIOD (B)

Target Dwell Time. Time spent looking at the burglar before establishing eye contact

Subject fixates on avatar to make eye contact (or on grey fixation point)

ANALYSIS PERIOD (C)

Premature Initiating Saccades. Frequency of initiating attempts before establishing eye contact

Avatar completes search (looks at 0-2 more houses) and makes eye (or fixation point turns green)

Subject must make a saccade to the burglar location

Feedback. Avatar responds to joint attention bid by gazing at the burglar's location (or fixation point blinks). Burglar appears behind bars

Figure 2. Schematic representation of trial sequence by condition. The eye symbol represents the fixation required by the participant and was not visible to the participant. Interest periods for each eye tracking analysis are indicated in red cells.

Search phase. Each trial began with a search phase. During this period, the participant and Alan were required to search for the burglar. The participant was instructed to search houses with blue doors (e.g., bottom row in Figure 1) and Alan searched houses with red doors (e.g., top row in Figure 1). Participants completed two blocks of trials. The location of the blue doors (i.e., top or bottom) was counterbalanced across blocks, and block order was counterbalanced across subjects. This controlled for effects of saccade trajectory on saccadic reaction times (SRTs).

Each time a participant fixated upon a blue door, it opened to reveal either an empty house or the burglar (Figure 2, first column). Participants were able to search their houses in any order they chose. On some trials, 1 or 2 blue doors were programmed to be already open, revealing an empty house. This forced participants to search their houses in a different order on different trials. Thus, the search pattern displayed by the avatar was realistically unpredictable.

Alan's behaviour was programmed so that he was still searching his allotted houses when the participants fixated back upon him. At this point he searched 0-2 more houses and then made eye contact. This ensured that, for a brief interval, participants were required to monitor Alan's gaze behaviour to determine whether he was ready to respond to, or initiate a joint attention bid. To further enhance the ecological validity of Alan's gaze behaviour, the onset latency of each eye movement was jittered with a uniform distribution between 500-1000 ms.

RJA. On RJA trials, the participant opened all the blue doors to find them empty (Figure 2, row 1). This meant that the burglar was hiding in one of Alan's houses. Once the participant fixated back on Alan's face, he searched 0-2 more houses before establishing mutual gaze with the participant. We randomised the location of the house that Alan searched last to ensure that it was not predictive of the burglar's location. Provided that the participant was still looking at Alan, he averted his gaze to guide the participant to the

burglar's location. In order to catch the burglar, the participant was required to follow Alan's gaze and fixate on the correct location. We refer to this eye movement as a "responding saccade".

IJA. On IJA trials, the participant found a burglar behind one of the blue doors (Figure 2, row 3). The relevant blue door "closed" to conceal the burglar once the participant fixated back on Alan's face. Again, Alan searched 0-2 more houses before making eye contact with the participant. Once eye contact was established, participants were required to initiate joint attention by fixating on the blue door that concealed the burglar. We refer to this eye movement as an "initiating saccade". Alan's gaze was programmed so that he always followed the participant's initiating saccade even if the participant fixated on the incorrect house. Importantly, Alan only responded to a participant's initiating saccade after eye contact had been established.

Feedback. For "correct" RJA and IJA trials, the burglar appeared behind bars to indicate that the participant and Alan had succeeded in achieving joint attention to capture the burglar (e.g., Figure 2, seventh column). On "incorrect" trials, the burglar appeared in red at its true location to provide negative feedback if participants (1) spent more than three seconds looking away from the target stimuli (e.g., fixating on the background, or Alan's allocated houses), (2) took longer than three seconds to execute a responding or initiating saccade after being guided (RJA trials) or establishing eye contact (IJA trials), or (3) made a responding or initiating saccade to an incorrect location. Furthermore, if participants took longer than three seconds to begin searching their houses at the beginning of the trial, red text reading "Failed Search" appeared on the screen and the trial was terminated.

Control conditions (RJAc and IJAc). We developed two conditions to control for the non-social task demands involved when responding to (RJAc; Figure 2, second row) and initiating (IJAc; Figure 2, fourth row) joint attention bids in the social conditions (i.e.,

task complexity, attentional load, and number of eye movements required). The only

difference between the control and social conditions was that the eyes of the avatar were

replaced by non-social cues (i.e., arrows). Specifically, (1) the virtual character's eyes

remained closed, (2) a grey fixation point was presented over the virtual character's nasion

until the participant completed their search and fixated it, (3) the fixation point turned

green when fixated (analogous to direct gaze on social trials), and (4) a green arrow

subtending three degrees of visual angle cued the burglar's location on RJAc trials

(analogous to averted gaze on RJA trials; see Caruana et al., 2015 for a video of example

trials from each condition).

Procedure

To ensure the testing environments of the two sites (UWA and MQ) were matched

as closely as possible, the same experimenter conducted every testing session, and all

participants were provided with the same instructions (see Supplementary Material 1).

Stimuli at both testing sites were presented at the same visual angle and eye movements

were recoded using an identical eye-tracking system and recording parameters (described

below).

Eye tracking. Eye-movements were recorded with a sampling rate of 500 Hz from

the right eye using a desktop-mounted EyeLink 1000 Remote Eye-Tracking System (SR

Research Ltd., Ontario, Canada). A chinrest was used to stabilise head movements and

standardise viewing distance. We conducted a 9-point calibration and validation at the

beginning of each block. Seven gaze-related areas of interest (GAOI) were used by our

gaze-contingent algorithm (depicted by blue rectangles in Figure 1). A GAOI covered each

of the six houses and the avatar. Eye movements were monitored online and recalibration

was conducted on trials where the participant made consecutive fixations on the borders or

outside the GAOIs. These recalibration trials were excluded from all analyses. On average

this accounted for 0.87% of trials from the HFA group ($SD = 1.14$) and 0.05% of trials

from the control group (*SD* = 0.22).

Joint attention task. The task was presented using *Experiment Builder* 1.10.165 (SR Research, 2004). At the beginning of the experiment, a scripted set of instructions was read aloud to the participant, and a series of cue cards were used to provide a schematic representation of the interactive eye-tracking interface (see Supplementary Materials 1). This supported the induction of the belief that participants were interacting with another person. Participants then completed two blocks of trials (Block 1 and Block 2). Each block comprised 27 trials from each condition (i.e., RJA, RJAc, IJA, IJAc). Social (RJA, IJA) and control (RJAc, IJAc) trials were presented in clusters of six trials throughout each block. These clusters began with a 1000 millisecond (ms) cue presented in the centre of the screen that read "Together" for a social cluster and "Alone" for a control cluster of trials. The randomisation of trial order within and across clusters was constrained to ensure that, within each block, conditions were matched on (1) the frequency that the burglar appeared in each location, (2) the number and location of houses that required searching on each trial, and (3) the number of gaze shifts made by the avatar before establishing eye contact.

There were four trial-order protocols that could be completed on each block. Two required the participant to search the upper row of houses (upper blocks), and two involved a search of the lower row of houses (lower blocks). For each pair of protocols, one began with a social cluster of trials (RJA, IJA) and the other began with a control cluster of trials (RJAc, IJAc). Each participant completed one upper and one lower protocol. Protocol and cluster order were counterbalanced across participants, and matched between the HFA and control groups.

Subjective task ratings. After the interactive game, participants completed the NASA Task Load Index (NASA-TLX; Hart, 2006), which is a standardised measure of subjective workload within human-machine environments. Participants completed the NASA-TLX for both the social and control conditions. This measure requires participants

to rate how demanding a task is on six dimensions, each on a continuous 20-point scale. The six subscales assessed the extent to which individuals found the tasks cognitively (Mental Demand) or physically (Physical Demand) challenging, or unduly fast-paced (Temporal Demand). It also measured the extent that individuals felt successful at the task (Performance), how hard they worked to achieve that level of performance (Effort), and how irritating or stressful (Frustration) they felt the task was.

In addition, participants rated the difficulty, naturalness, intuitiveness, and pleasantness of the social and control conditions on a 5-point Likert scale. We also asked them to indicate whether they preferred playing the game together with Alan (i.e., social condition) or alone (i.e., control condition). They also rated this preference on a continuous 10-point scale where 0 indicated a complete preference for the social interaction, and 10 indicated a complete preference for completing the task alone. To determine whether participants believed that they were interacting with another human, we asked them to rate how cooperative they thought Alan was on a 5-point scale (5 = extremely cooperative). This provided them with an opportunity to declare whether they believed (or not) that Alan was a real person. Participants were then debriefed about the true nature of the interaction, and were asked to use the same 5-point scale to rate how convinced they were that the virtual character was controlled by a real person (5 = completely convinced).

Scores

Accuracy. For each condition (i.e., RJA, RJAc, IJA, IJAc) we measured accuracy as the proportion of trials (excluding recalibration trials) where the participant succeeded in catching the burglar.

Responding to joint attention. For correct RJA and RJAc trials, we also measured *saccadic reaction time* – the latency (in ms) between the presentation of the orienting cue (gaze for RJA, arrow for RJAc) and the onset of the responding saccade that resulted in a fixation at the correct burglar location (see Figure 2, Analysis Period A).

Initiating joint attention. For correct IJA and IJAc trials, we derived two measures of participants' readiness to establish eye contact before initiating joint attention. First we considered the *target dwell time* – the total amount of time (in ms) that the participant fixated on the burglar from the time it had been found until they fixated back on Alan's face (see Figure 2, Analysis Period B). Second, we determined the proportion of trials containing *premature initiating saccades* whereby participants made a saccade from Alan's face to the burglar location before Alan had established eye contact (IJA) or the fixation point had turned green (IJAc; see Figure 2, Analysis Period C).

Subjective task ratings. We calculated difference scores for each individual by subtracting ratings in the control condition from the social condition.

Statistical Analyses

Parametric tests were used for all measures of task performance, including Analysis of Variance (ANOVA) and Pearson r Correlation Coefficients. A Levene Test for Equal Variance was used to assess whether each data set was equivariant between groups. Test statistics were adjusted for lack of equivariance where necessary. Non-parametric tests (Mann-Whitney U Test) were used on all measures of subjective experience, consistent with Svensson's (2001) recommendations for the analysis of ratings data. We used two-tailed tests with *p*-values of .05 for ANOVAs and Mann-Whitney U tests. Since we had clear directions of prediction for the Pearson r Correlation Coefficients, we used 1-tailed significance tests. Outliers of distance and influence were identified using SPSS Casewise Diagnostics and Cook's Distance estimates. Significant outliers were removed from correlation analyses where appropriate.

Parametric analysis. The parametric analyses of scores for task accuracy, saccadic reaction time (RJA), target dwell time (IJA), and premature initiating saccades (IJA) each comprised three main steps. First, we conducted an analysis of variance (ANOVA). Group (HFA versus control) was a between-subjects factor and condition (social versus control)

and block (Block 1 versus Block 2) were within-subjects factors. Second, any significant group*condition interaction effects were followed-up using paired t-tests that assessed the effect of condition in each group, and independent samples t-tests that assessed the effects of group for each condition. Third, we followed-up significant three-way interaction effects that suggested a learning effect (i.e., improvement in Block 2) that was more pronounced for one group (e.g., HFA) and one condition (e.g., social). We did this first by using separate 2 (group) x 2 (condition) ANOVAs for Block 1 and Block 2 data. Then we used Pearson r Correlation Coefficients to determine if higher SQ scores (indicating high systemisers) were associated with poorer scores in the relevant measure in Block 1, and if higher SQ scores were associated with greater gains in the relevant performance scores from Block 1 to Block 2, in each group.

Nonparametric tests. We used Mann-Whitney U two-sample tests to determine if the subjective-rating difference scores of the two groups differed significantly using a *p* value of .05.

Results

Summary statistics for all dependant variables are shown in Table 2. For each analysis, we report main effects of condition and group, and group*condition and group*condition*block interaction effects (see Supplementary Material 2 for a full summary of the ANOVA output).

Table 2.

Means and Standard Deviations for all Outcome Measures by Group, Condition and Block

	HFA		Control	
	RJA	RJAc	RJA	RJAc
Accuracy				
Block 1	72.08	89.88	90.01	93.66
	(23.17)	(10.87)	(8.17)	(10.62)
Block 2	80.83	87.44	96.08	93.90
	(24.12)	(12.83)	(4.98)	(7.17)
Saccadic Reaction Time				
Block 1	782.41	395.69	544.82	350.54
	(397.43)	(153.87)	(131.64)	(56.68)
Block 2	553.75	351.59	513.37	324.96
	(145.30)	(135.91)	(124.57)	(45.63)
	IJA	IJAc	IJA	IJAc
Accuracy				
Block 1	98.02	94.03	98.68	99.13
	(3.49)	(11.18)	(2.27)	(2.08)
Block 2	97.57	96.67	100.00	100.00
	(4.53)	(8.20)	(0.00)	(0.00)
Target Dwell Time				
Block 1	1637.57	902.18	972.92	774.40
	(1465.81)	(319.06)	(463.66)	(173.10)
Block 2	1053.64	733.40	761.96	694.24
	(840.44)	(266.53)	(217.46)	(128.28)
Premature Initiating Saccades				
Block 1	8.12	3.88	5.82	3.35
	(7.04)	(3.06)	(4.77)	(4.00)
Block 2	5.65	1.53	4.71	0.94
	(6.29)	(1.87)	(6.01)	(1.30)

Responding to Joint Attention

Accuracy. Accuracy data is shown in Figure 3. Participants made significantly

more errors on RJA trials than RJAc trials (main effect of condition, $F(1,32) = 6.64$, $p =$

.015, $\eta_p^2 = 0.17$). Individuals with HFA made significantly more errors than the control

group (main effect of group, $F(1,32) = 9.06$, $p = .005$, $\eta_p^2 = 0.22$). There was a significant

group*condition interaction ($F(1,32) = 5.22$, $p = .029$, $\eta_p^2 = 0.14$). Posthoc tests revealed

that the HFA group made significantly more errors than the control group on RJA trials

($t(17) = -3.05$, $p = .007$) but not on RJAc trials ($t(32) = -1.65$, $p = .108$). There was no

significant group*condition*block effect ($F(1,32) = 0.95$, $p = .338$, $\eta_p^2 = 0.03$), suggesting

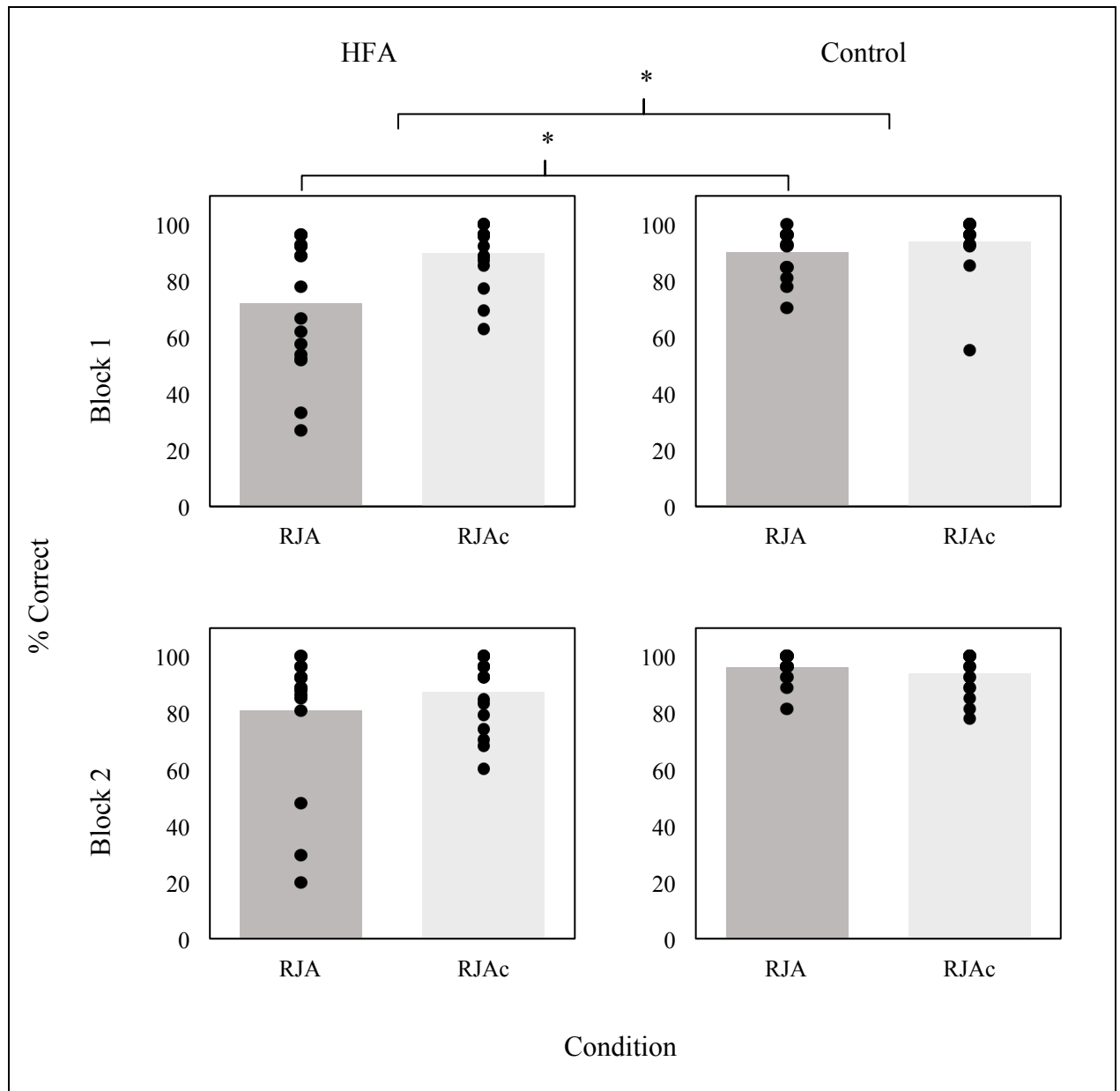that the group*condition difference was consistent across blocks.

Figure 3. Percentage of correct trials in RJA and RJAc conditions, separated by block and group.

Bar graphs represent group means. Data points represent individual means.* denotes significant differences

between groups and conditions.

Saccadic reaction time. Individual data points and group means are presented in Figure 4 by group, condition, and block. Participants were significantly slower to respond on RJA trials than RJAc trials ($F(1,32) = 86.74$, $p < .005$, $\eta_p^2 = 0.73$). The main effect of group ($F(1,32) = 3.67$, $p = .064$, $\eta_p^2 = 0.10$) and the group*condition interaction ($F(1,32) = 3.91$, $p = .057$, $\eta_p^2 = 0.11$) failed to reach significance. However, there was a significant group*condition*block interaction ($F(1,32) = 4.49$, $p = .042$, $\eta_p^2 = 0.12$) indicating different group*condition effects in the two blocks. For Block 1, there was a significant group*condition interaction ($F(1,32) = 5.09$, $p = .031$, $\eta_p^2 = 0.14$), with the HFA group being significantly slower than controls to respond on RJA trials ($t(19.47) = 2.34$, $p = .030$) but not on RJAc trials ($t(20.26) = 1.34$, $p = .269$). For Block 2, there was no significant group*condition interaction ($F(1,32) = 0.11$, $p = .743$, $\eta_p^2 < 0.01$). Two-way ANOVAs in each group separately (factors: condition and block) revealed differential learning effects in each group. There was a significant condition*block interaction in the HFA group ($F(1,14) = 5.45$, $p = .033$, $\eta_p^2 = 0.25$) due to a significant improvement in SRTs between blocks on RJA ($t(16) = 2.80$, $p = .013$) but not RJAc trials ($t(16) = 1.87$, $p = .080$). There was no significant condition*block interaction or main effect of block in the control group (all ps > 0.15). This suggested that improvement in SRTs was unique to the HFA group in the RJA condition.
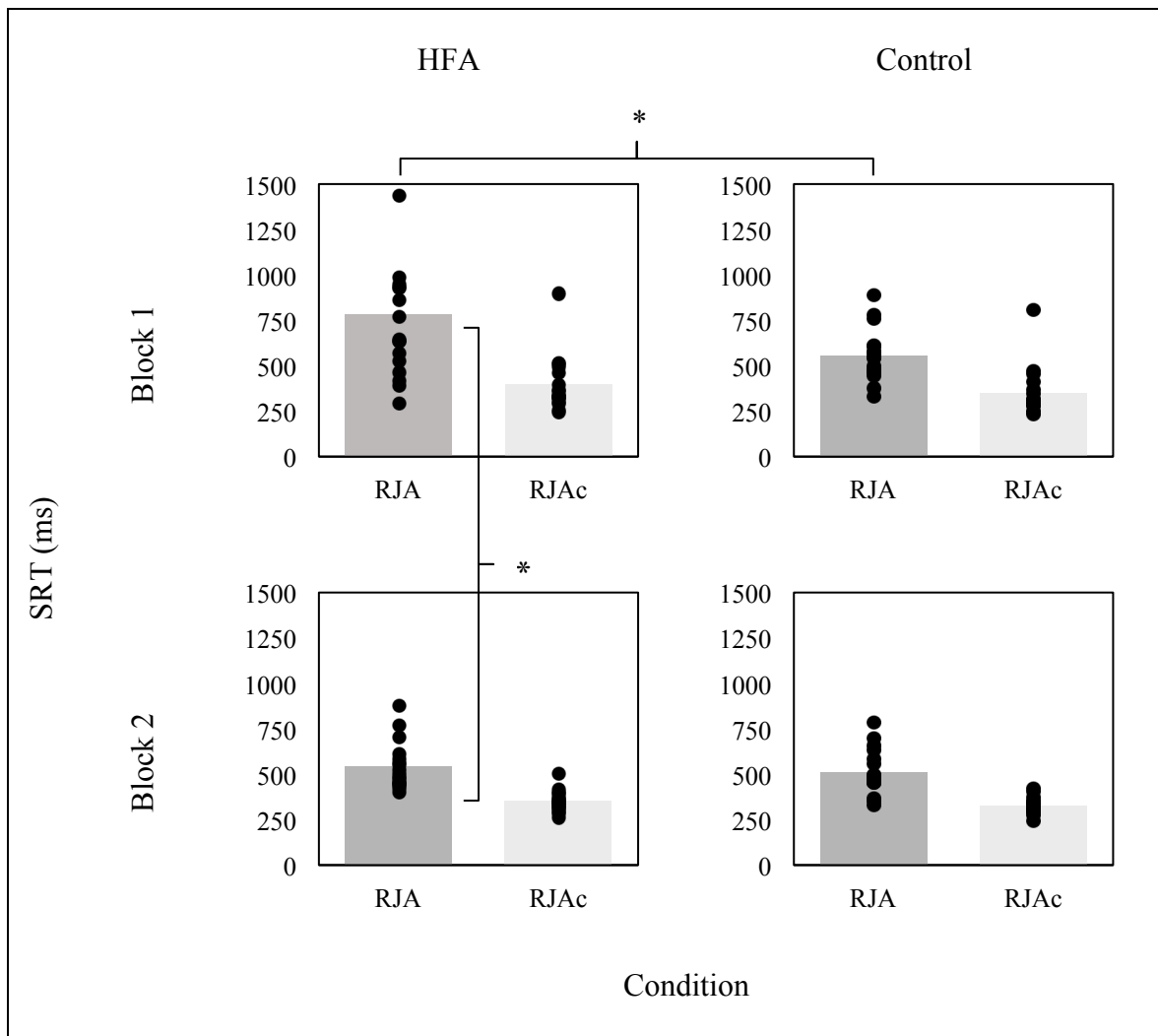
Figure 4. Average saccadic reaction time in milliseconds for each block, separated by condition and group. Bar graphs represent group means. Data points represent individual means. * denotes significant differences between groups and blocks.

The significant three-way interaction (i.e., group*condition*block) for SRTs suggested that, relative to controls, participants with HFA improved most in the social condition over the course of the experiment. Could individual differences in this improvement be related to individual differences in systemising? As illustrated in Figure 5, there were moderate and positive Pearson r correlation coefficients between SQ and SRTs in the RJA ($r = 0.49$, $p = .026$, 1-tailed) and RJAc ($r = .51$, $p = .021$, 1-tailed) conditions for the HFA group during Block 1 (one outlier excluded), indicating that people with higher SQ scores were initially slower to respond in both conditions. There were no

significant correlations between SQ and SRTs in either condition in Block 2 for the HFA group, or in any condition or block for the control group (all ps > .141).

To investigate whether improvement across blocks was related to individual differences in SQ scores, we assessed the correlation between SQ and SRT improvement scores for each individual (i.e., Block 1 minus Block 2 SRTs). For the HFA group, SQ was positively associated with SRT improvement on RJA trials ($r = 0.66$, $p = .003$, 1-tailed) and RJAc trials ($r = 0.43$, $p = .049$, 1-tailed). It was not significant for either condition in the control group (all ps > 0.17). To assess the specificity of this effect, we carried out the same correlational analyses with AQ, EQ, and non-verbal IQ. None of these variables showed a significant positive correlation with task performance gains in the HFA group (all ps > .056). Thus, within the HFA group, it appeared that high systemisers were initially the slowest to respond to gaze and arrow cues in Block 1. However, they also demonstrated the greatest improvement, becoming more similar to low systemisers with HFA and controls by Block 2 (see Figure 5).
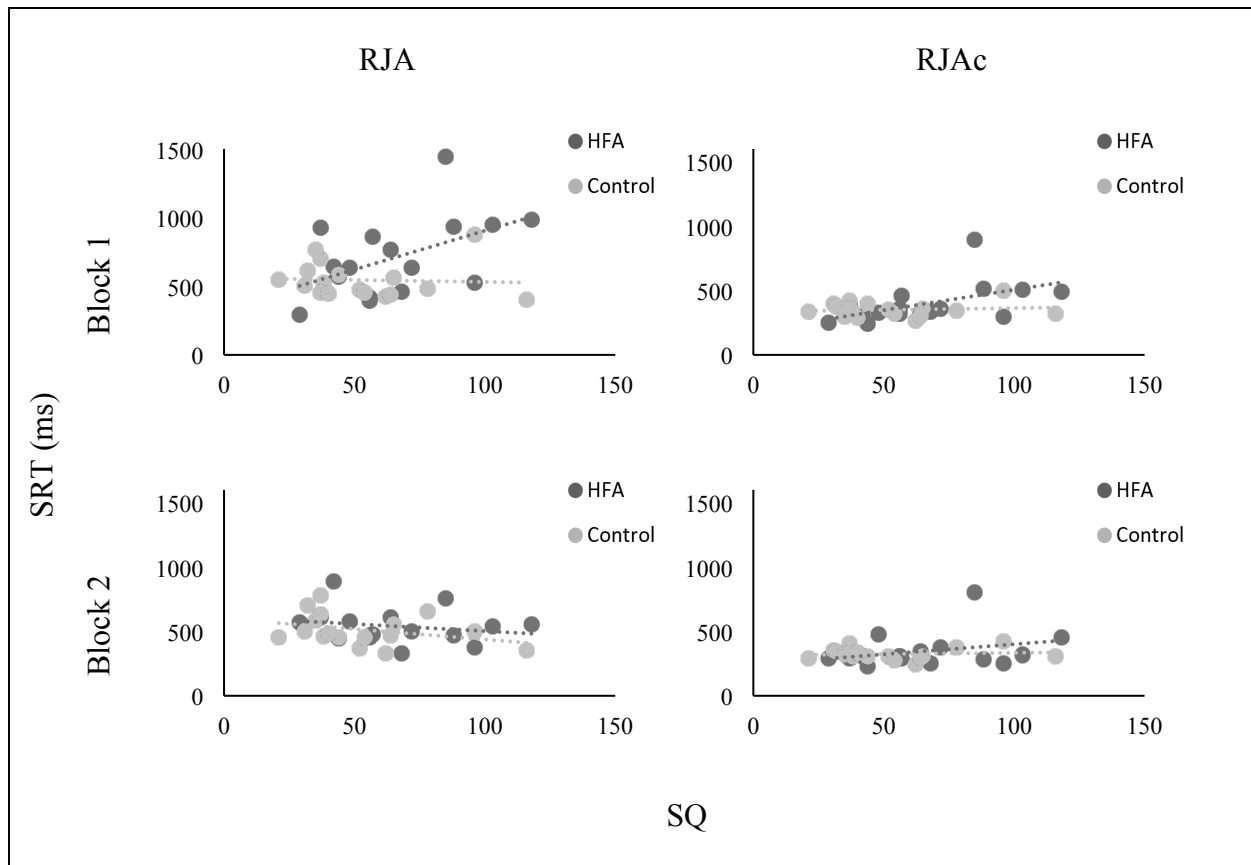
Figure 5. Scatter plots illustrating the relationship between scores on the Systemising Quotient (SQ) and saccadic

reaction times. Data from each condition and block are plotted separately. Data from the HFA and control groups

are plotted on the same graphs for comparison.

Initiating Joint Attention

Accuracy. Accuracy data is shown in Figure 6. There was no significant main effect of condition ($F(1,32) = 1.50$, $p = .230$, $\eta_p^2 = 0.45$). However, individuals in the HFA group made significantly more errors than the control group ($F(1,32) = 7.04$, $p = .012$, $\eta_p^2 = 0.18$). There was no group*condition interaction ($F(1,32) = 2.17$, $p = .150$, $\eta_p^2 = 0.06$) because participants with HFA were less accurate than controls in both the IJA ($t(32) = -2.37$, $p = 0.24$) and IJAc ($t(16.64) = -2.26$, $p = .037$) conditions. There was also no group*condition*block interaction ($F(1,32) = 1.51$, $p = .227$, $\eta_p^2 = 0.05$).

Controls and some individuals with HFA performed at ceiling for both IJA and IJAc in Block 2 (see Figure 6) meaning that we may have underestimated the true size of the group effect. Whilst not ideal, this is not overly concerning since the group effect was nevertheless significant. It also did not cause a spurious group*condition interaction, since this was non-significant. Further, since the HFA group did not significantly differ in accuracy for IJA and IJAc ($t(16) = 1.33$, $p = 0.20$) any true group*condition interaction that we missed would have been driven by a difference between IJA and IJAc in controls, which is not of interest in this study.
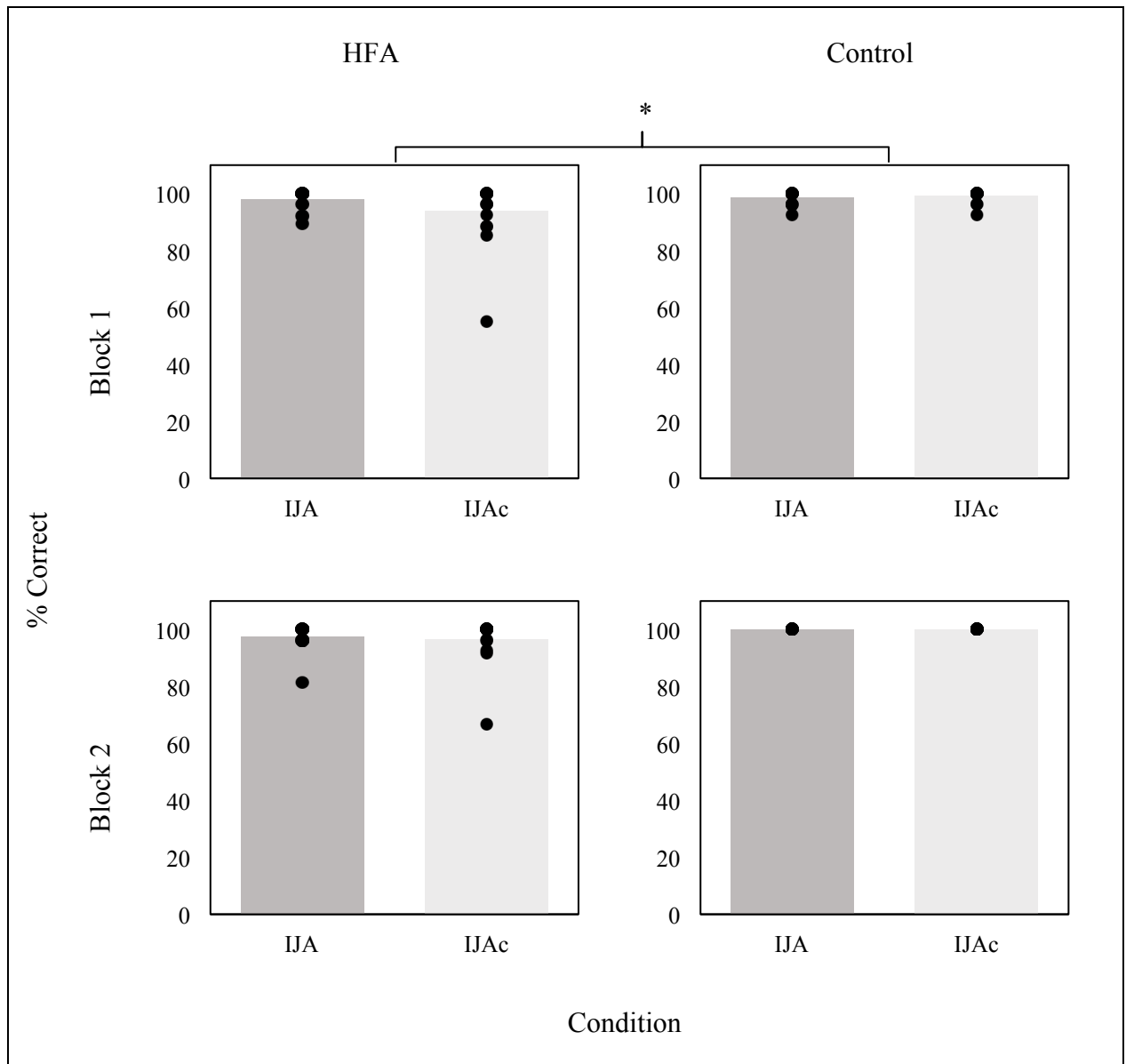
Figure 6. Percentage of correct trials in IJA and IJAc conditions, separated by block and group.

Bar graphs represent group means. Data points represent individual means. * denotes significant differences

between groups.

Target dwell time. Individual data points and group means are presented in Figure 7. Participants spent significantly more time fixated on the burglar before establishing eye contact on IJA trials relative to analogous eye movements on IJAc trials ($F(1,32) = 7.73$, $p = .009$, $\eta_p^2 = 0.20$). There was no main effect of group, ($F(1,32) = 2.73$, $p = .108$, $\eta_p^2 = 0.08$), group*condition interaction ($F(1,32) = 2.76$, $p = .107$, $\eta_p^2 = 0.08$), or group*condition*block interaction ($F(1,32) = 2.05$, $p = .162$, $\eta_p^2 = 0.06$). As depicted in Figure 7, dwell times tended to be longest during IJA trials in Block 1.
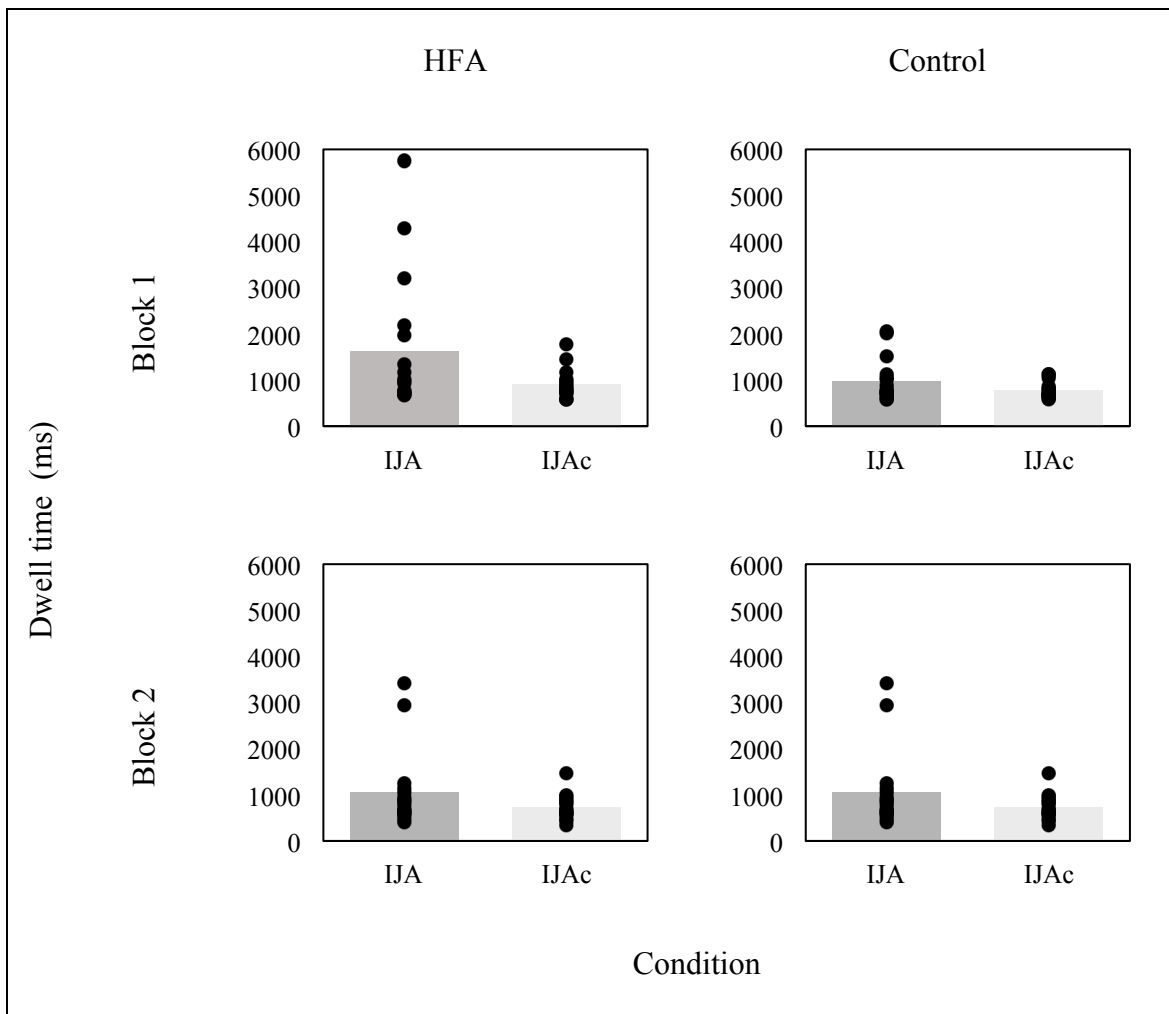
Figure 7. Average dwell times on the burglar before looking back to the avatar to establish eye contact (IJA) or looking back to the fixation point (IJAc). Data is displayed separately for each block, condition and group. Bar graphs represent group means. Data points represent individual means.

Premature initiating saccades. Individual data points and group means are presented in Figure 8. Participants made significantly more premature attempts at initiating joint attention on IJA trials relative to analogous eye movements on IJAc trials ($F(1,32) = 19.85$, $p < .005$, $\eta_p^2 = 0.38$). There was no significant main effect of group, ($F(1,32) = 0.86$, $p = .360$, $\eta_p^2 = 0.03$), group*condition interaction ($F(1,32) = 0.42$, $p = .522$, $\eta_p^2 = 0.01$), or group*condition*block interaction ($F(1,32) = 0.50$, $p = .486$, $\eta_p^2 = 0.02$).
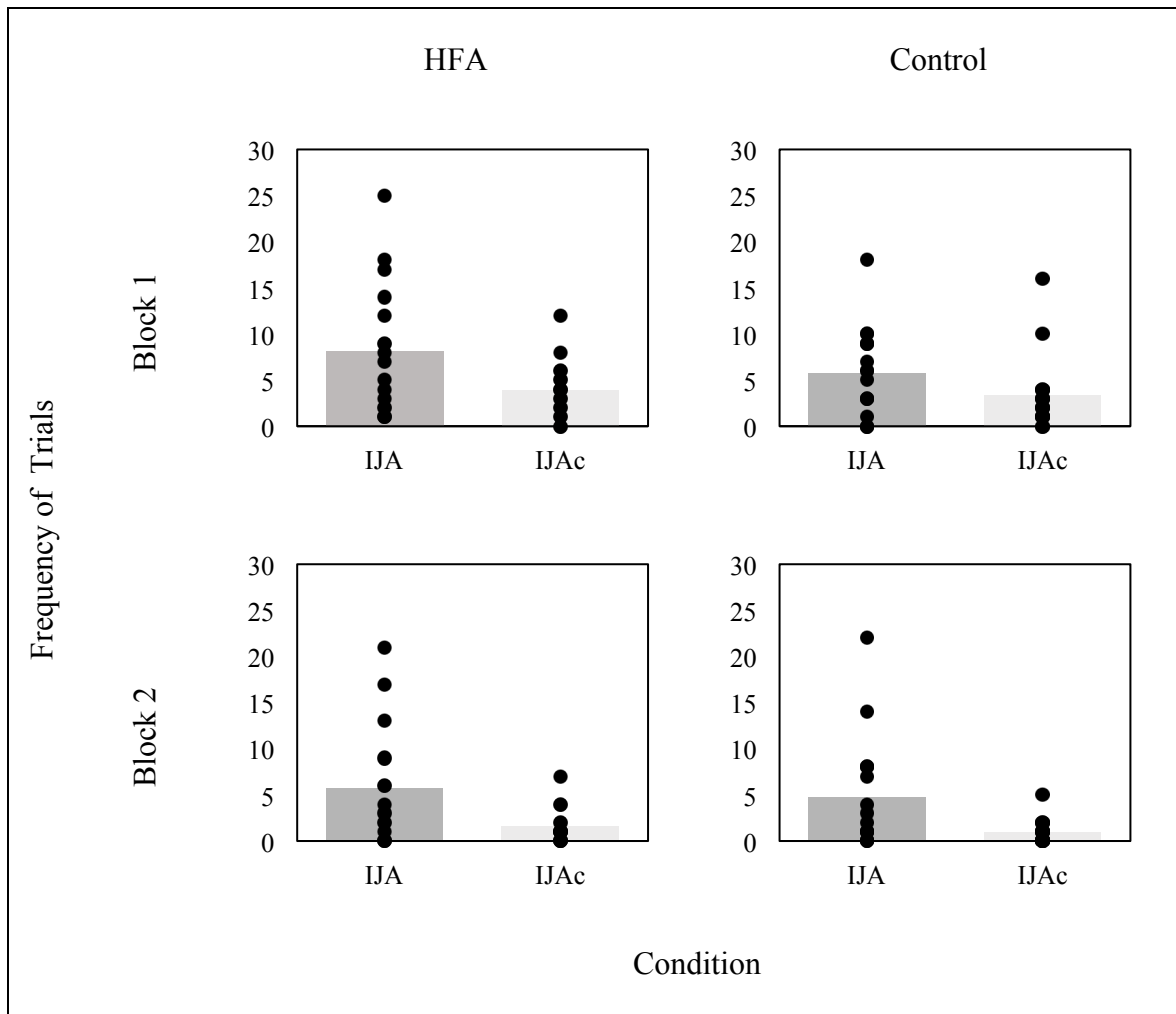
Figure 8. Average frequency of trials containing a saccade from the central stimulus to the burglar before the avatar makes eye contact (IJA) or the fixation point turns green (IJAc). Data is displayed separately for each block, condition and group. Bar graphs represent group means. Data points represent individual means.

Subjective Task Ratings

NASA-TLX. We measured the extent of perceived task load in the social (RJA, IJA) and control (RJAc, IJAc) conditions. Summary statistics are reported in Table 3. Effect scores were calculated for each participant and subscale by subtracting the raw ratings on the control condition from those on the social condition tasks. There were no significant differences in task load effects between the HFA and control groups on any of the NASA-TLX subscales using these effect scores (all $ps > .067$).

Experience ratings. There were no significant differences between the HFA and control groups on ratings of task difficulty, naturalness, intuitiveness, or pleasantness (all $ps > .322$). Summary statistics are reported in Table 3.

Partner's cooperativeness. Prior to debrief, the virtual character was rated as being significantly less cooperative by individuals in the HFA group ($M = 4.47$, $SD = 0.62$) than those in the control group ($M = 4.94$, $SD = 0.24$; $U = 84.50$, $p = .038$).

Belief that the virtual character was controlled by a real person. With the exception of two participants in the HFA group, all participants said that they were completely convinced that the virtual character was controlled by a real person. These two participants made no objection about the true nature of the interaction until they were debrief, even though they were previously asked to make a number of judgements about their partner. In fact, these participants provided comments that suggested our deception was successful. For instance, when asked whether they preferred the social condition or the control condition, one participant commented "*Together task* [i.e., social condition]. *More accurate because you can see the other person's perspective.*" Overall, there was no significant difference between the HFA ($M = 4.24$, $SD = 1.15$) and control groups ($M = 4.88$, $SD = 0.33$) on self-ratings of their conviction in the genuine nature of the interaction ($U = 91.50$, $p = .067$).

Preference for social condition over control condition. The HFA group reported significantly stronger preferences for playing the task alone ($M = 6.18$, $SD = 3.05$) compared to the control group ($M = 4.00$, $SD = 2.21$; $U = 81.50$, $p = .029$).

Table 3.

Summary Statistics for Subjective Task Ratings by Scale, Subscale and Group

	HFA			Control			Statistics	
	Social	Control	Effect	Social	Control	Effect	U	p
NASA-TLX								
Mental	9.12 (6.06)	7.35 (6.18)	1.76 (6.30)	5.24 (4.66)	5.00 (4.32)	0.24 (3.05)	119.00	.394
Physical	7.24 (5.25)	7.00 (6.19)	0.24 (3.95)	4.53 (3.74)	3.94 (3.78)	0.59 (1.70)	122.50	.454
Temporal	10.06 (6.44)	8.35 (6.90)	1.71 (6.17)	8.71 (4.93)	6.65 (5.73)	2.06 (4.56)	136.50	.786
Performance	5.32 (3.90)	5.18 (4.99)	0.18 (3.94)	4.76 (3.17)	3.47(2.94)	1.29 (0.85)	91.50	.067
Effort	10.65 (5.51)	7.71 (6.68)	2.94 (6.06)	6.88 (5.07)	5.35 (4.96)	1.53 (3.06)	127.00	.563
Frustration	6.35 (6.77)	4.12 (5.82)	2.24 (6.38)	2.82 (2.77)	2.12 (2.23)	0.71 (0.99)	130.50	.634
Experience								
Difficulty	2.47(1.23)	1.53 (0.94)	0.94 (1.56)	1.47 (0.62)	1.06 (0.24)	0.41 (0.62)	120.00	.413
Naturalness	3.76 (1.09)	4.00 (1.06)	-0.24 (1.44)	3.94 (0.75)	4.47 (0.72)	-0.53 (0.62)	115.00	.322
Intuitiveness	3.65 (1.27)	3.94 (1.20)	-0.29 (0.92)	4.41 (0.71)	4.65 (0.49)	-0.24 (0.56)	126.50	.540
Pleasantness	3.35 (1.32)	3.41 (1.23)	-0.06 (0.75)	4.47 (0.72)	4.65 (0.61)	-0.18 (0.53)	125.50	.518

Note. Ratings for each subscale on the NASA-TLX are provided on a continuous 20-point scale. Experience ratings were provided on a 5-point scale (1=low, 5=high). Summary statistics are provided in the format of *M (SD)*. Group differences in the effect of condition were tested using a two-sample Mann-Whitney U test.

148

Discussion

Difficulty establishing joint attention, whether by responding to another person or directing their attention, is a cardinal feature of autism (American Psychiatric Association, 2013). However, to date, experimental paradigms investigating aspects of joint attention in autism have often failed to capture the intentional nature of joint attention as it occurs in the context of natural social interactions. In the current study, we addressed this issue by means of a novel virtual reality paradigm in which the joint attention episode was preceded by a search phase. This provided an ecologically valid context in which participants had to determine whether to respond to or initiate joint attention bids. Having determined their social role, participants then had to wait for eye contact with an onscreen virtual partner before either following or guiding him to the target location. Although the virtual character was controlled by a computer, participants were successfully deceived into thinking that they were interacting with a real person. This is the first study of its kind in autism research and demonstrates the potential of such an approach for investigating joint attention and social interactions more generally in this population. This study also revealed some intriguing differences in responding and initiating joint attention behaviours between participants with and without autism.

Responding to Joint Attention Bids

Compared to typically-developed adults, individuals with HFA made significantly more errors when attempting to respond to their partner's joint attention bid, but did not significantly differ in their accuracy when responding to arrow cues in the non-social control condition. Analysis of eye movements also demonstrated that individuals with HFA were significantly slower than controls in executing saccades in response to their partner's joint attention bid, at least during the first half of testing. Again, these group differences were specific to the social condition: SRTs in response to the arrow cue in the control condition did not significantly differ between groups. The lack of group differences

on the control task suggests that difficulties in responding to joint attention bids are independent of impairments in oculomotor control, attention orienting, and executive function, given that the social and control conditions presented equivalent demands on these cognitive abilities.

In order to understand why individuals with HFA performed poorer, specifically on social trials, we must consider how the social (RJA) and control (RJAc) conditions differed. Aside from the obvious perceptual differences, the gaze cues in the social condition were more ambiguous than the arrow cues in the control condition. This is because participants had to determine whether their partner was averting his gaze to search for the burglar, or to initiate joint attention towards the burglar's location. In this context, the participant can disambiguate averted gaze as a communicative bid for joint attention if it is preceded by eye contact, rather than averted gaze in a different location. In contrast, the solitary arrow cue in the control condition is completely unambiguous. Therefore, it is possible that individuals with HFA made more errors when attempting to respond to joint attention bids in the social condition, and were initially slower when they did succeed, because they had difficulties processing the ostensive information conveyed when their partner made eye contact with them. This is consistent with previous findings using a multi-face gaze-cueing paradigm in which adults with HFA – unlike typically-developed peers – did not use eye contact to evaluate the relevance of an upcoming gaze shift (Böckler et al., 2014).

Interestingly, this interpretation also aligns with the subjective reports of participants with HFA. For example, one person reported: *"I felt a bit anxious during the together task* [i.e., social condition]. *The alone task* [i.e., control condition] *was easier because it was clear what the dot and arrow meant."* Another participant said that she was specifically looking for a "formula" to understand the social interaction. She explained that she learned during the control condition that she had to look back at the central stimulus.

She then applied this rule in the social condition, and that made it easier for her to look at Alan's gaze.

These subjective reports were also consistent with the finding that participants with HFA had longer SRTs in the first block of testing but not in the second. This suggests that they learned to efficiently respond to joint attention bids once they had developed an understanding of the rules governing the social interaction. For example, one participant said:

> *I don't normally look at peoples' eyes. You think I'm looking you in the eye now?* [looks at experimenter]*, but I'm focusing on an imaginary point a bit in front of your face. In the game I had to look at the eyes. . . then I thought, "Why are eyes harder than arrows?" So I decided to treat the eyes like arrows.*

One implicit rule that participants had to learn was that an averted gaze shift that followed eye contact communicated the location of the burglar, whereas a gaze shift that did not follow eye contact was not socially communicative. Within the HFA group, it was the participants who had the greatest tendency to understand complex phenomena as rule-governed systems (i.e., high systemisers) who started with the poorest performance and made the greatest improvements from the first to second blocks of testing. One explanation for this effect is that some people with HFA may have greater difficulties understanding social signals, and the rules of social interactions, and thus develop systemising tendencies as a compensatory mechanism. Since this is the first study to examine the relationship between individual differences in joint attention and systemising, it is important that the reliability of this relationship is tested by future studies.

It would also be interesting if future studies used similar virtual reality paradigms to investigate some of the conscious systemising strategies that individuals with autism use during social interactions. Understanding the types of strategies that facilitate and hinder social information processing and communication may inform intervention research by

elucidating the techniques that may assist individuals with autism to (1) effectively understand and communicate with others, and (2) make social interactions more pleasant, and less intimidating.

Initiating Joint Attention Bids

The current study is the first to provide an objective measure of initiating joint attention using eye movement analyses. Again, participants with HFA made more errors than control participants. However, in contrast to the responding condition, this was not specific to the social version of the task, with significant group differences found for both IJA and IJAc conditions. Participants made a combination of errors on IJA and IJAc trials that may be explained by difficulties in oculomotor control or attention. Specifically, they (1) took longer than three seconds to make a saccade towards the burglar location after establishing eye contact (IJA) or the fixation point turned green (IJAc), (2) initiated joint attention to an incorrect location by fixating another house first, or (3) spent more than three seconds fixating on task-irrelevant portions of the screen (e.g., the background, or on houses allotted to their partner). Since these errors occurred in both IJA and IJAc conditions, it is possible that individuals with autism may have subtle deficits in oculomotor control or attention that prevent them from fixating on task-relevant stimuli in a timely fashion.

The lack of group differences in accuracy for the IJA condition was further corroborated by the fact that individuals with HFA did not demonstrate significantly longer target dwell times, or more premature initiating saccades, than typically-developed peers. Nevertheless, the variability of these behaviours within the HFA sample was high. This was particularly true for the target dwell time analysis. Approximately one third of the HFA sample was 2-4 times slower than controls to make eye contact after finding the burglar (see Figure 7). These delays in establishing eye contact may be of clinical significance when considered in the context of real-life, fast-paced interactions. In order to

adequately account for the high level of individual variability in initiating behaviours, future work is needed in which these paradigms are applied across larger samples, with individuals across the autism spectrum, and at different stages of development. This will help elucidate the constellation of individual characteristics that are associated with difficulties and strengths in initiating joint attention. It would also be useful if future studies could obtain additional measures of individuals' social functioning in daily life situations in order to determine whether variability in task performance is related to daily social functioning.

Compared to measures of target dwell time, we found less variation within the HFA group in our analysis of premature initiating saccades. Interestingly, we found that individuals from both groups were more likely to prematurely saccade from the central stimulus to the burglar on IJA trials (i.e., before the avatar made eye contact) than on IJAc trials (i.e., before fixation point turned green). This may have occurred because participants expected their social partner to follow them even when they did not wait to establish eye contact (i.e., the social condition), whereas it was clear that this would not happen when they were interacting with the computer (i.e., the control condition). Future studies could manipulate participants' beliefs about the human agency of the virtual character. This would determine whether participants make fewer premature initiating saccades when they believe that the virtual character is controlled by a computer.

Subjective Task Ratings

In addition to the objective measures of performance on the burglar task, we also asked participants for their subjective impression of the task. We found no significant differences between groups on subjective ratings of task load, difficulty, naturalness, intuitiveness, or pleasantness. However, compared to typically-developed individuals, participants with HFA rated their partner as being significantly less cooperative. This may have been driven by the fact that individuals with HFA made more errors than controls,

and attributed these errors to their partner's behaviour rather than their own.

Individuals with HFA also provided significantly stronger preference ratings for the control conditions than the social conditions. The subjective comments of many individuals with HFA suggested that the most unpleasant aspect of the social interaction was the need to establish and use eye contact. For example, participants with HFA expressed that (1) *"...the eyes were harder to figure out"*; (2) *"Alone* [i.e., control condition] *was easier to complete because you didn't have to catch his eye to tell him where to go"*, and (3) *"When they* [i.e., eyes] *were closed I didn't have to worry about him and what he wants. Didn't have to have the patience to wait for him."* These comments highlight that the individuals with HFA were aware that establishing eye contact and using gaze as a communicative technique was challenging for them. This is consistent with a larger body of literature that suggests that individuals with autism find it difficult to use eye contact to understand and regulate social interactions (e.g., Pelphrey, Shultz, Hudac, & Vander Wyk, 2011; Senju & Johnson, 2009).

Some individuals with HFA also indicated that, while they preferred to complete the task alone than with Alan, they also preferred the virtual interaction than real-life face-to-face interactions. They offered two reasons for this. First, some individuals suggested that the virtual interface provided them with an opportunity to engage in a social interaction with less anxiety than they often experience in social settings. For example, one participant said: *"I don't like dealing with people so this was better. Feels like you're socialising, but not . . .* [I felt] *more relaxed"*. Another participant explained that she preferred real-life interactions over the virtual interaction, but only if eye contact could be avoided. Otherwise, she would prefer virtual reality: *"*[Virtual reality] *makes it more comfortable . . . I am an 'audio' person. I like to ask things if they're not clear. So I would prefer real life. Not face-to-face, but side-by-side"*.

Second, some adults with HFA suggested that the virtual reality paradigm allowed

them to focus on specific aspects of their social interaction with Alan without being overwhelmed by multiple cues. For example, one participant with HFA explained that the virtual interface acted as a "filter" which made the social interaction more accessible to them because they only had to focus on one aspect of their partner at a time (i.e., their eyes): *"Easier to segment the task and interaction in virtual reality. Only focus on one thing"*. Another person said: *"I can interact but don't have too many things to think about"*. These insights suggest that our virtual reality paradigm might be a useful training tool for interventions focusing on social information processing and communication in people with HFA. It has the degree of experimental control required to gradually accustom individuals to one aspect of social communication at a time (e.g., eye gaze) using stimuli that may be less intimidating than real people. For example, it could be used to gradually expose individuals to virtual partners who are increasingly unpredictable by manipulating aspects of their gaze behaviour (e.g., duration of gaze shifts, degree of cooperativeness) as well as the addition of other social cues (e.g., ostensive eye-brow flashes, gestures and speech).

Conclusion

This is the first study to use an ecologically valid, objective, quantified, and experimentally controlled measure to test both RJA and IJA in individuals with autism. Our data suggests that adults with HFA demonstrate significant difficulties in responding to and initiating joint attention bids. However, initiating difficulties were not consistent across the sample of participants with HFA. These findings encourage further work investigating the individual characteristics that may account for the heterogeneity of joint attention abilities in autism. Our analysis suggests that systemising tendencies may be a possible candidate.

Beyond these specific findings, this study also highlights the potential for interactive virtual reality paradigms in guiding the training of social information

processing and communication skills among individuals with autism. Specifically, virtual

reality paradigms can be used to identify the precise aspects of social interactions that

individuals with autism find difficult, and the strategies that are likely to make social

communication more effective and less stressful. Equipping individuals with the skills to

use eye gaze to regulate joint attention interactions has the potential to improve

opportunities for social learning, language development and relationship regulation

(Adamson et al., 2009; Baron-Cohen, 1995; Charman, 2003; Mundy et al., 1990; Murray et

al., 2008; Tomasello, 1995).

References

Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. A. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of Autism & Developmental Disorders, 39*(1), 84-96. doi: 10.1007/s10803-008-0601-7

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5 ed.). Washington, DC

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interactions. *Child Development, 55*(4), 1278. doi: 10.1111/1467-8624.ep7302943

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5-17.

Baron-Cohen, S. (2008). Autism, hypersystemizing, and truth. *Quarterly Journal of Experimental Psychology, 61*(1), 64-75. doi: 10.1080/17470210701508749

Bates, E., Benigni, L., Bretherton, I., Csmaioni, L., & Volterra, V. (1979). *The emergence of symbols: cognition and communication in infancy.* New York: Academic Press.

Böckler, A., Timmermans, B., Sebanz, N., Vogeley, K., & Schilbach, L. (2014). Effects of observing eye contact on gaze following in high-functioning autism. *Journal of Autism & Developmental Disorders, 44*(7), 1651-1658. doi: 10.1007/s10803-014-2038-5

Brock, J. (2011). Commentary: Complementary approaches to the developmental cognitive neuroscience of autism – reflections on Pelphrey et al. (2011). *Journal of Child Psychology and Psychiatry, 52*(6), 645-646. doi: 10.1111/j.1469-7610.2011.02414.x

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation & Developmental Disabilities Research Reviews, 10*(3), 169-175. doi: 10.1002/mrdd.20036

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

Cary, M. S. (1978). The role of gaze in the initiation of conversation. *Social Psychology, 41*(3), 269-271. doi: 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. doi: 10.1098/rstb.2002.1199

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., & Drew, A. (1997). Infants with autism: An investigation of empathy, pretend play, joint attention, and imitation. *Developmental Psychology, 33*(5), 781-789. doi: 10.1037/0012-1649.33.5.781

Chawarska, K., Klin, A., & Volkmar, F. (2003). Automatic attention cueing through eye movement in 2-year-old children with autism. *Child Development, 74*(4), 1108-1122.

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. *Developmental Psychology, 40*(2), 271-283. doi: http://dx.doi.org/10.1037/0012-1649.40.2.271

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. doi: 10.1037/0033-2909.133.4.694

Hart, S. G. (2006). *NASA-Task Load Index (NASA-TLX); 20 Years Later.* Paper presented at the Human Factors and Ergonomics Society 50th Annual Meeting, Santa Monica: HFES.

Hobson, J., & Hobson, P. (2007). Identification: The missing link between joint attention and imitation? *Development and Psychopathology, 19*(02), 411-431. doi: doi:10.1017/S0954579407070204

Johnson, M. H., Griffin, R., Csibra, G., Halit, H., Farroni, T., De Haan, M., Tucker, L. A., Baron-Cohen, S., & Richards, J. (2005). The emergence of the social brain network: Evidence from typical and atypical development. *Development and Psychopathology, 17*(03), 599-619. doi: doi:10.1017/S0954579405050297

Kylliainen, A., & Hietanen, J. K. (2004). Attention orienting by another's gaze direction in children with autism. *Journal of Child Psychology and Psychiatry, 45*(3), 435-444. doi: 10.1111/j.1469-7610.2004.00235.x

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. *Journal of Autism and Developmental Disorders, 30*(3), 205–223. doi: 10.1023/A:1005592401947

Loveland, K. A., & Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *Journal of Autism & Developmental Disorders, 16*(3), 335-349. doi: 10.1007/BF01531663

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20*(1), 115-128. doi: 10.1007/bf02206861

Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. *Development and Psychopathology, 6*(03), 389-401. doi: doi:10.1017/S0954579400006003

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: The contribution of non-verbal communication measures. *Journal of Child Psychology and Psychiatry, 27*(5), 657-669. doi: 10.1111/j.1469-7610.1986.tb00190.x

Murray, D. S., Creaghead, N. A., Manning-Courtney, P., Shear, P. K., Bean, J., & Prendeville, J. A. (2008). The relationship between joint attention and language in children with autism spectrum disorders. *Focus on Autism & Other Developmental Disabilities, 23*(1), 5-14. doi: 10.1177/1088357607311443

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? Is it automatic? Is it social? *Development and Psychopathology, 20*(01), 79-97. doi: 10.1017/S0954579408000047

Okada, T., Sato, W., Murai, T., Kubota, Y., & Toichi, M. (2003). Eye gaze triggers visuospatial attentional shift in individuals with autism. *Psychologia, 46*(4), 246-254. doi: 10.2117/psysoc.2003.246

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders, 24*(3), 247-257. doi: 10.1007/bf02172225

Osterling, J., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology, 14*(02), 239-251. doi: doi:10.1017/S0954579402002031

Pelphrey, K. A., Shultz, S., Hudac, C. M., & Vander Wyk, B. C. (2011). Research Review: Constraining heterogeneity: the social brain and its development in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 52*(6), 631-644. doi: 10.1111/j.1469-7610.2010.02349.x

Phillips, W., Baron-Cohen, S., & Rutter, M. (1992). The role of eye contact in goal detection: Evidence from normal infants and children with autism or mental handicap. *Development and Psychopathology, 4*(03), 375-383. doi: doi:10.1017/S0954579400000845

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

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research, 24*(3), 715-718. doi: 10.1016/j.cogbrainres.2005.02.007

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature, 253*(5489), 265-266. doi:10.1038/253265a0

Senju, A., & Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences, 13*(3), 127-134. doi: http://dx.doi.org/10.1016/j.tics.2008.11.009

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome and developmental delays. *Monographs of the Society for Research in Child Development, 64*(1), 1-114. doi: 10.1111/1540-5834.00002

SR Research. (2004). Experiment Builder (Version 1.10.165). Ontario.

Stone, W. L., Ousley, O. Y., & Littleford, C. D. (1997). Motor imitation in young children with autism: what's the object? *Journal of Abnormal Child Psychology, 25*, 475-485. doi: 10.1023/A:1022685731726

Svensson, E. (2001). Guidelines to statistical evaluation of data from rating scales and
    questionnaires. *Journal of Rehabilitation Medicine, 33*(1), 47-48.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham
    (Eds.), *Joint Attention: Its Origins and Role in Development*. Hillsdale: Lawrence
    Erlbaum Associates.

Vlamings, P. H. J. M., Stauder, J. E. A., van Son, I. A. M., Mottron, & Laurent. (2005).
    Atypical visual orienting to gaze- and arrow-cues in adults with high functioning
    autism. *Journal of Autism and Developmental Disorders, 35*(3), 267-277. doi:
    10.1007/s10803-005-3289-y

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., Weil,
    L., &Wakabayashi, A. (2006). Predicting Autism Spectrum Quotient (AQ) from the
    Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain
    Research, 1079*(1), 47-56. doi: http://dx.doi.org/10.1016/j.brainres.2006.01.012

Wong, C., & Kasari, C. (2012). Play and joint attention of children with autism in the
    preschool special education classroom. *Journal of Autism and Developmental
    Disorders, 42*(10), 2152-2161. doi: 10.1007/s10803-012-1467-2

Chapter 5

The Neural Time Course of Evaluating Self-initiated

Joint Attention Bids

Nathan **Caruana**[123], Peter **de Lissa**[124] & Genevieve **McArthur**[12]

[1] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[2]ARC Centre of Excellence in Cognition and its Disorders

[3]Perception in Action Research Centre

[4] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

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 350ms 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.

The Neural Time Course of Evaluating Self-initiated Joint Attention Bids

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.

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 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).

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.

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 et al., 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.

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 directing-attention 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 (cf. 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.

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 reported by 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 (i.e., 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 to 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 movemenets of another human via live infrared eye-tracking. In reality, the virtual agent was controlled by a gaze-contingent algorithm (cf. 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 (i.e., "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).

Method

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

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 written consent before participating.

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 represented 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.

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).

Stimuli

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 x 12 degrees of visual angle in the centre of a computer screen (a 60 x 34 cm Samsung SynchMaster SA950 HD LED monitor 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 degrees of visual angle, with 15 degrees 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).

    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 (cf. Caruana et al., 2015). A fixation point subtending 1.4 degrees of visual angle was positioned over the nose. Green arrow stimuli, which protruded from this fixation point replaced the gaze stimuli, and subtended 4 degrees of visual angle (see Figure 1 for a comparison of social and control task stimuli).
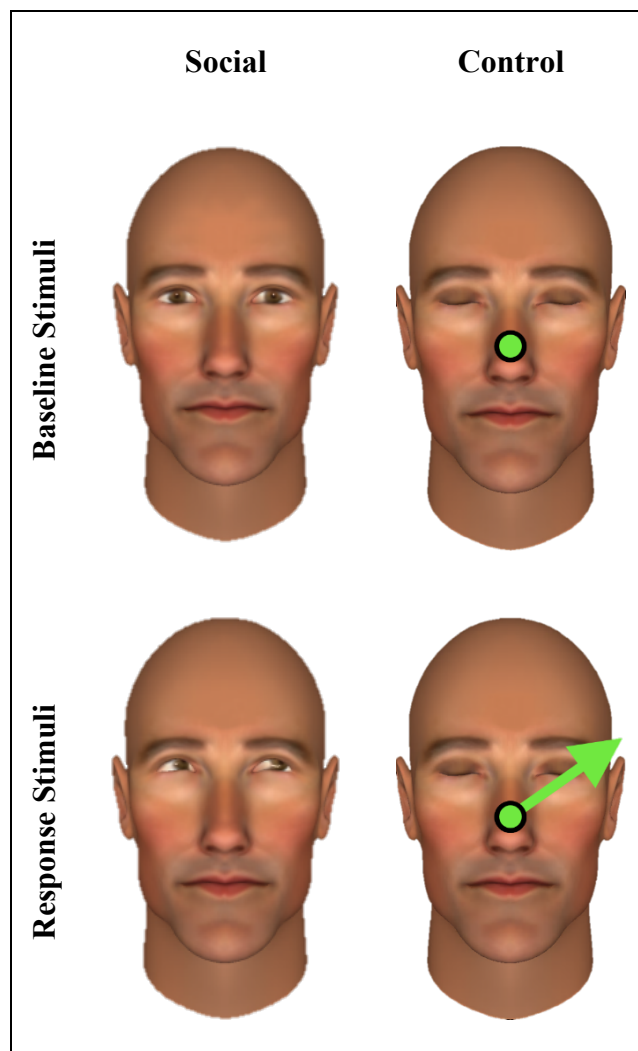

Figure 1. Central stimuli used in the social and control versions of the gaze-contingent task.

Stimulus Conditions

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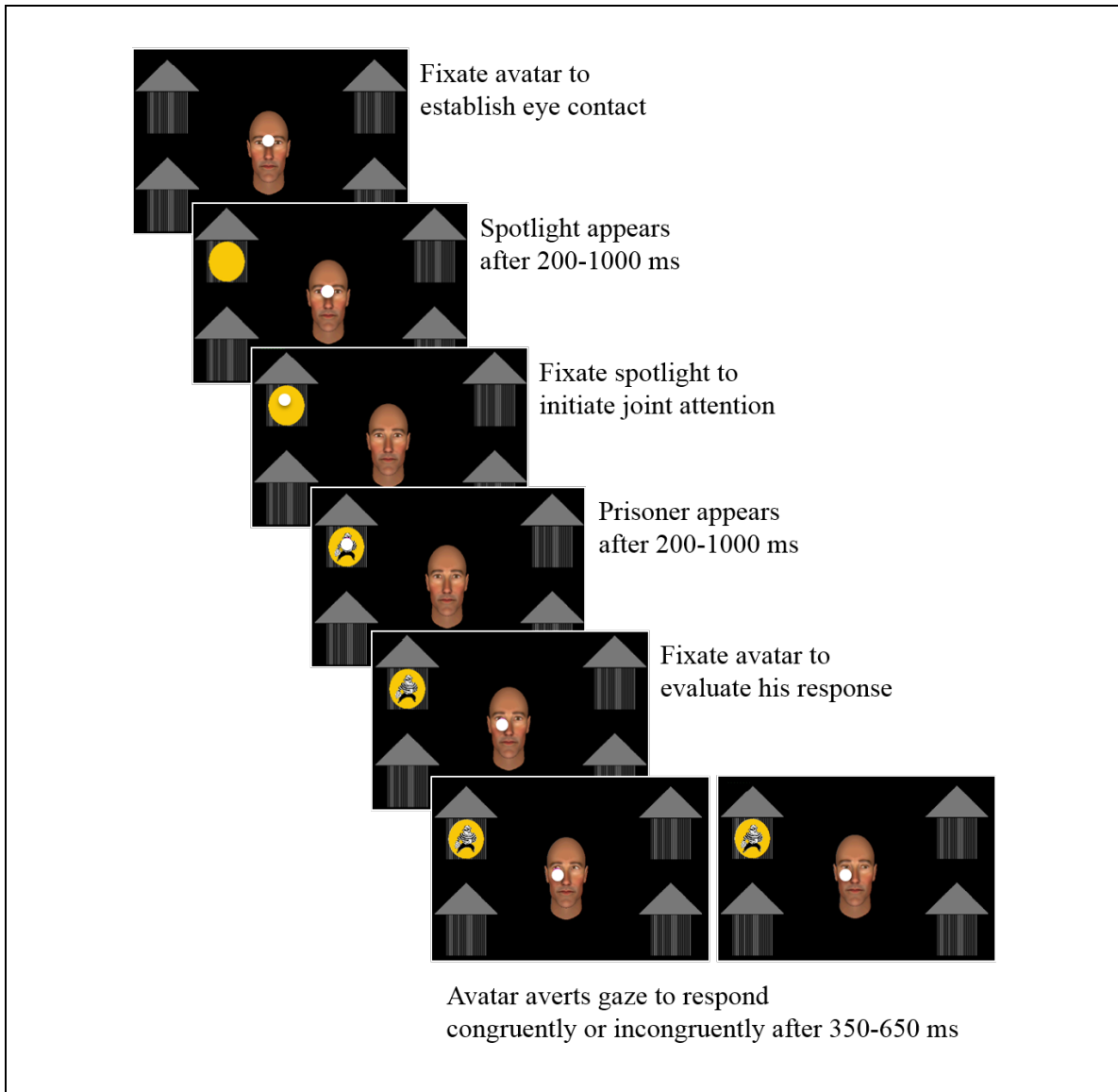
Figure 2. Schematic representation of trial sequence. White circle represents the location of the participant's gaze and was not part of the stimuli visible to the participant.

A prisoner then attempted to escape from one of four buildings that were each located in a different corner of the display screen. 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 counterbalanced across all acquisition blocks, and thus the virtual character was equally likely to gaze towards one of the three houses not 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.

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 (e.g., 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.

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 kΩ. 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.

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.

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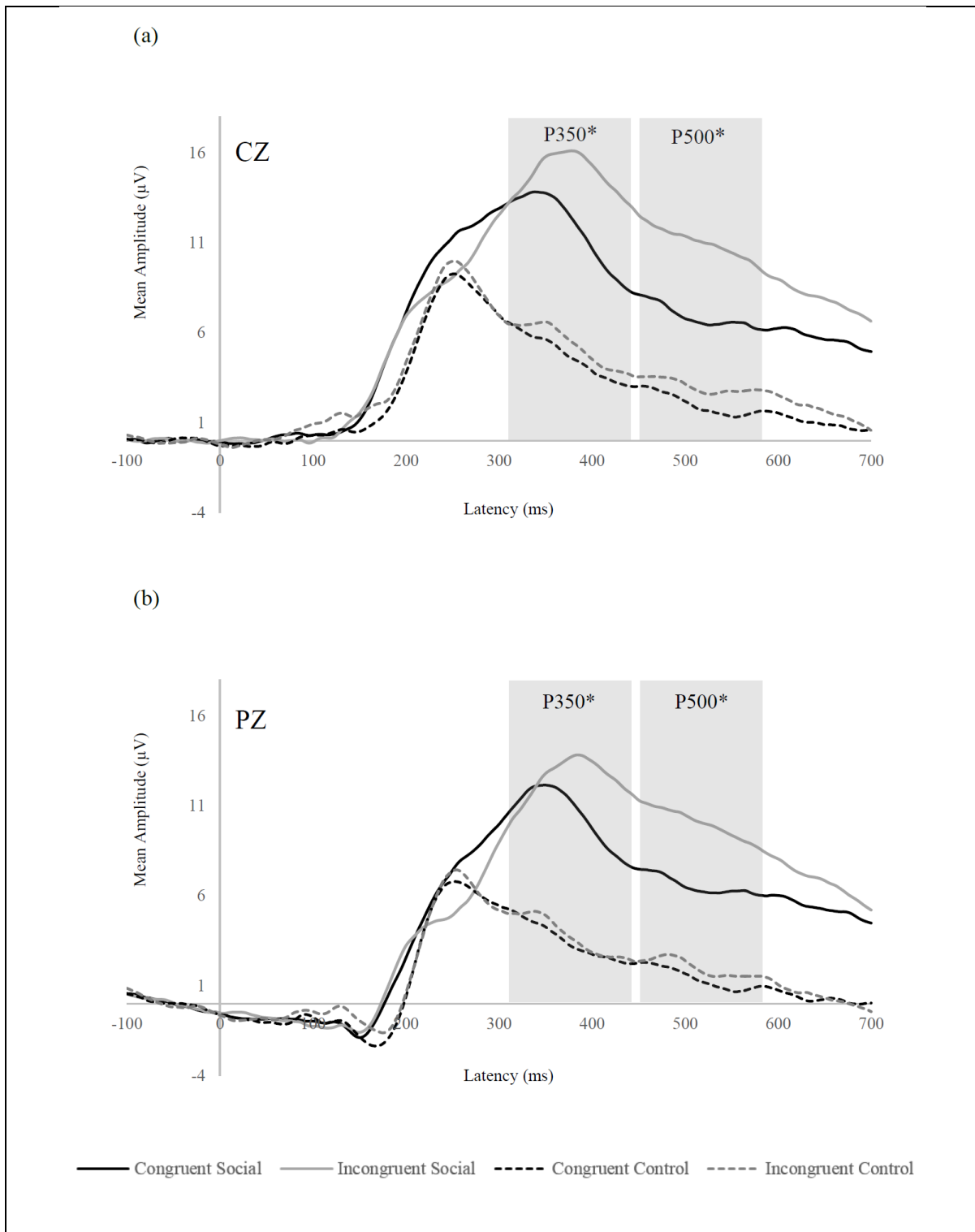
Figure 3. Group average waveforms comprising the P350 and P500 at (a) Cz and (b) Pz electrodes. Epochs were time-locked to the onset of the virtual character's gaze shift. * indicates a significant Group*Condition interaction effect.

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, *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 (310-440ms) 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 (310-440ms) and P500 (from 450-580ms), both at CZ and PZ, and the N170 response recorded at P7 and P8 (107-237ms). 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 in the P500 interval).

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).

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.

Results

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 $ps > 0.29$).

Table 1

Post-experimental Inquiry Ratings of Subjective Experience

Task Aspect	Social	Control
	M(SD)	*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).

Social condition. Participants in the social condition rated the interaction as feeling "moderately" to "very" natural on average ($M = 2.58$, $SD = 1.26$). 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.

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.

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 provided 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).

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 Figure 3, and for the N170 at P7 and P8 are shown in Figure 4. Topographies demonstrating the effect of stimulus condition (congruent – incongruent) are depicted separately for the social and control conditions in Figure 5.

Table 2.

Summary Statistics for Amplitude and Latency Measures by Electrode

	CZ		PZ	
	Congruent	**Incongruent**	**Congruent**	**Incongruent**
P350 mean amplitude				
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)
P500 mean amplitude				
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**
N170 peak amplitude				
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)*.

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 ($t(18) = 0.64$, $p = .532$).
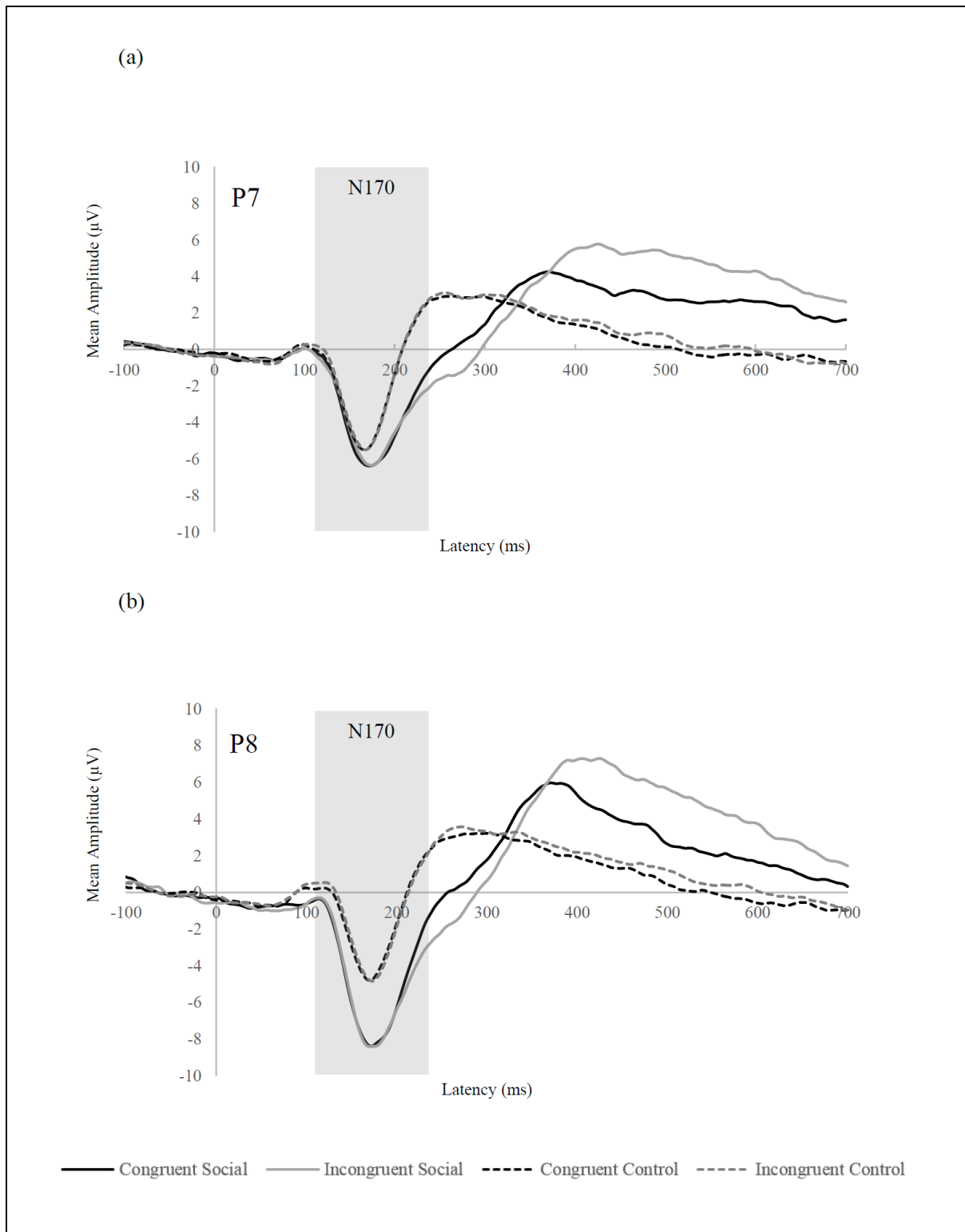
Figure 4. Group average waveforms comprising the N170 at (a) P7 and (b) P8 electrodes. Epochs were time-locked to the onset of the virtual character's gaze shift.
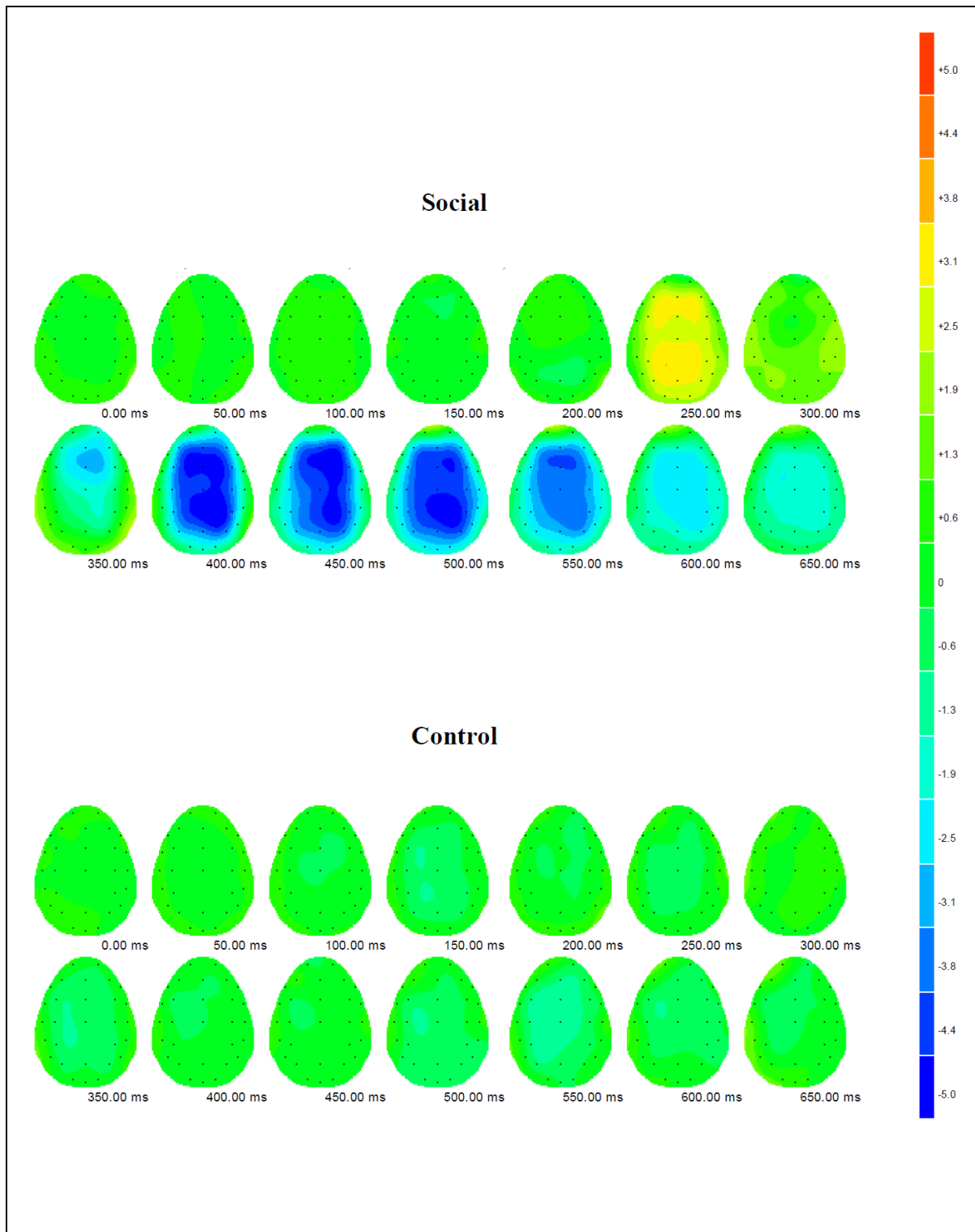
Figure 5. Effect topographies (congruent – incongruent) by group condition.

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.

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).

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 the 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 Figure 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' 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 Figure 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 congruency 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 & Zang, 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 200-300 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 congruency 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) reported that increased activity was 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 that reinforces future instances of social 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 cues 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 information processing and 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. doi: 10.1162/jocn.1996.8.6.551

Bruner, J. S. (1974). From communication to language, A psychological perspective. *Cognition, 3*(3), 255-287. doi: 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. doi: 10.1097/WNR.0b013e3280ebb44b

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

Cary, M. S. (1978). The role of gaze in the initiation of conversation. *Social Psychology, 41*(3), 269-271. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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, 8*(6), 544-54 1-11. doi: 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. doi: 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*(0), 123-136. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1016/j.clinph.2005.07.005

Kimball, S., & Mattis, P. (1995). GIMP Image Manipulation Program (Version 2.8.2). Berkley.

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological Bulletin, 100*(1), 78-100. doi: http://dx.doi.org/10.1037/0033-2909.100.1.78

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. doi: 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 Science, 3*(6), 207-216. doi:10.1016/S1364-6613(99)01329-7

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

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., & Vogeley, K. (2014). Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *NeuroImage, 101*, 124-137. doi: 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*. doi: 10.1093/scan/nsq068

Puce, A., Smith, A., & Allison, T. (2000). ERPs evoked by viewing facial movements. *Cognitive Neuropsychology, 17*(1), 221-239. doi: 10.1080/026432900380580.

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. doi: http://dx.doi.org/10.1016/j.bandc.2003.04.002

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36,* 441- 462. doi: 10.1017/S0140525X12002452

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah N. J., Fink, G. R., & Vogeley, K. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience, 22*(12), 2702-2715. doi: 10.1162/jocn.2009.21401.

Schweinberger, S. R., Kloth, N., & Jenkins, R. (2007). Are you looking at me? Neural correlates of gaze adaptation. *Neuroreport, 18*(7), 693-696. doi: 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). doi: 10.3389/fnint.2010.00005

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. doi: 10.1111/1469-8986.00083

Watanabe, S., Miki, K., & Kakigi, R. (2002). Gaze direction affects face perception in humans. *Neuroscience Letters, 325*(3), 163-166. doi:10.1016/S0304-3940(02)00257-4

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. doi: 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. doi: 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. doi: 10.1371/journal.pone.0094339

Chapter 6

Beliefs About Human Agency Influence Gaze-related ERPs During Joint Attention

Nathan **Caruana**[123], Peter **de Lissa**[124] & Genevieve **McArthur**[12]

[1] Department of Cognitive Science, MACQUARIE UNIVERSITY, Sydney

[2] ARC Centre of Excellence in Cognition and its Disorders

[3] Perception in Action Research Centre

[4] Department of Psychology, MACQUARIE UNIVERSITY, Sydney

Abstract

Background: Humans use the eye movements of others during social interactions to understand their intentions and perspective. Unfortunately, most studies that have investigated the neural processing of gaze have used non-interactive paradigms that lack ecological validity. Consequently, little is known about the factors that influence the neural processing of gaze in ecologically valid interactive contexts.

Method: The current study measured adults' P350 and N170 ERPs whilst they interacted with a character in a virtual reality paradigm. Some participants believed the character was controlled by a human ("avatar" condition, n=19); others believed it was controlled by a computer program ("agent" condition, n=19). In each trial, participants initiated joint attention in order to direct the virtual character's gaze towards a target. In 50% of trials the character gazed towards the target (congruent responses), and in 50% of trials the charactered gazed to a different location (incongruent response).

Results: In the avatar condition, the character's incongruent gaze responses generated a significantly larger P350 peak at centro-parietal sites than congruent gaze responses. In the agent condition, the P350 effect was strikingly absent. Left occipitotemporal N170 responses were significantly smaller in the agent condition compared to the avatar condition for both congruent and incongruent gaze shifts.

Discussion: These data suggest that the belief that a virtual character represents a human may recruit mechanisms that discriminate the social outcome of a gaze shift after approximately 350 ms. These mechanisms may also modulate the early perceptual processing of gaze as indexed by a larger left occipitotemporal N170. These findings in turn suggest that the ecologically valid measurement of social cognition depends on paradigms that simulate social interactions with real human beings.

Beliefs About Human Agency Influence Gaze-related ERPs During Joint Attention

Introduction

Humans are skilled in extracting social signals conveyed by another's gaze during interactions. We use gaze to understand the emotions and intentions of others, and to coordinate joint attention experiences with them (i.e., a common focus of attention). Joint attention is a pivotal cognitive ability as it supports the development of language and social learning skills (Adamson, Bakeman, Deckner, & Romski, 2009; Baron-Cohen, 1995; Charman, 2003; Mundy, Sullivan, & Mastergeorge, 2009; Murray et al., 2008; Tomasello, 1995). Impairments in the ability to initiate and respond to bids for joint attention also constitute one of the most homogenous characteristics of autism (Pelphrey, Shultz, Hudac, & Vander Wyk, 2011). It is therefore important to understand the neural processing of gaze during joint attention interactions.

Unfortunately, progress in understanding the processing of gaze during joint attention has been hindered by the challenge of developing neurophysiological paradigms that can simulate an ecologically valid interaction whilst simultaneously ensuring tight experimental control (Schilbach et al., 2013). Virtual reality has been the solution adopted by many social neuroscientists using functional magnetic resonance imaging (fMRI; Caruana, Brock, & Woolgar, 2015; Pfeiffer et al., 2014; Schilbach et al., 2010) and event-related potentials (ERP; Caruana, de Lissa, & McArthur, 2015) because virtual characters can be fully controlled and realistically convey anthropomorphic form and behaviour (see Georgescu, Kuzmanovic, Roth, Bente, & Vogeley, 2014 for review).

The increased use of virtual characters in social neuroscience raises an important empirical and methodological question: *Is it important for participants to believe that a virtual character is being controlled by a real human?* This is typically achieved by deceiving participants, which introduces practical and ethical issues into an experiment. In order to determine if such issues are justified (i.e., to inform best practice), the current

study investigated the effect that beliefs of human agency have on gaze-related neural processes during joint attention interactions with a virtual partner.

The Influence of Agency Beliefs on Subjective Experience

Agency beliefs refer to the extent to which an individual believes a stimulus to represent the online behaviour and intentions of another human. Virtual reality studies have begun to distinguish between virtual characters that are believed to be computer-controlled "agents" or human-controlled "avatars". This distinction was first made by von der Pütten, Krämer, Gratch, and Kang (2010) who investigated the influence of agency beliefs on subjective experience and self-disclosure during one-way conversational interactions with agents and avatars. In this paradigm, participants were asked to respond to a series of questions asked by a virtual character. The presence of contingent head-nods made by the virtual character resulted in an increase in participants' self-disclosure, and a reduction of low dominance negative feelings (e.g., weak, shy) measured using the Positive and Negative Affect Schedule (Watson, Tellegen, & Clark, 1988). Participants also reported less negative low-dominance feelings (e.g., scared, ashamed) when they believed the virtual character to be an avatar rather than an agent. Nevertheless, the authors concluded that the participants' experience with the virtual character was mostly influenced by the perceived realism of the stimulus rather than beliefs about whether the character was controlled by another human. This interpretation seems at odds with the data since (1) the belief of whether the virtual character was an avatar or agent was found to influence subjective experience on one outcome measure, and (2) perceived realism was measured offline after the virtual interaction was over using self-disclosure, which is heavily influenced by personality traits (e.g., extraversion; Peter, Valkenburg, & Schouten, 2005). Furthermore, these data do not elucidate whether agency beliefs influence gaze-related neural processes during joint attention interactions.

The Influence of Agency Beliefs on Gaze Processing

Pfeiffer et al. (2014) were the first to provide evidence that human agency beliefs influence the neural processing of social stimuli during virtual interactions. In their fMRI study, they asked participants to interact with an anthropomorphic virtual character. Participants were instructed that in some trials the virtual character's gaze would be controlled by a computer program, and in other trials it would be controlled by another human using the online recordings of their eye movements. On each trial, participants initiated a joint attention bid by fixating on one of two squares located on either side of the virtual character's face. The virtual character averted his gaze to look at the same square (a congruent response that achieved joint attention) or at the alternate square (an incongruent response that avoided joint attention). Each block comprised five trials. The "congruency" of each block was manipulated by adjusting the proportion of congruent trials and incongruent trials in each block. At the end of each block, participants decided whether they believed the virtual character was an avatar or agent. The authors reported that participants were more likely to believe that the avatar was being controlled by a human in blocks where the virtual character responded congruently more often to achieve joint attention. Blocks in which participants believed that they interacted with a human were associated with increased activation of the ventral striatum relative to blocks in which participants believed that they were interacting with a computer-controlled agent. This effect has been interpreted as reflecting the hedonic experience of achieving a self-initiated joint attention bid (Schilbach et al., 2010). However, given that beliefs about human agency were clearly influenced by the congruency of the virtual character's response, this cannot be interpreted as direct evidence for the influence of human agency beliefs on the neural processing of gaze during joint attention interactions.

More direct evidence for the effect of human agency beliefs on the neural processing of gaze comes from two ERP studies. Wykowska, Wiese, Prosser, and Müller,

(2014) asked participants to complete a gaze-cueing task in which they were presented with a pair of eyes imbedded in a robot face. On each trial, participants were asked to use a button box to identify the location of a target presented to the left or right side of the robot's face. The target was preceded by a valid gaze cue (i.e., the robot shifted its gaze toward the target location) or an invalid gaze cue (i.e., the robot looked in the opposite direction to the target location). On some trials, participants were instructed that the robot was controlled by a human; on other trials, they were instructed that the robot was pre-programmed.

P1 ERP responses were measured at posterior-occipital sites 100-140 ms after the onset of the target. The P1 is a positive ERP peak that is believed to reflect neural processes associated with visual attention (Itier & Taylor, 2004). The authors reported that P1 responses to targets were significantly larger following the presentation of valid gaze cues than invalid gaze cues. However, this effect was only observed when participants believed the robot to be controlled by a human, and not when they believed it to be pre-programmed. This outcome is particularly striking given that (1) agency beliefs were only manipulated via instruction (i.e., the tasks were identical), and (2) this instruction was irrelevant to the task. The authors explained this effect within the Intentional Stance Model of Social Cognition, suggesting that the perception of agency recruits the neural mechanisms that support mentalising (i.e., the cognitive ability to understand the mental states of others). These mentalising mechanisms are argued to have a top down "sensory gain" effects on visual processes (Wykowska, et al., 2014).

While Wykowska et al.'s (2014) findings certainly support the idea that agency beliefs influence gaze-related effects on visual attention, they do not elucidate whether agency beliefs influence the perceptual processing of gaze shifts specifically. The P1 was time-locked to the presentation of the target that appeared a long time (i.e., 600 ms) after the presentation of the gaze cue. Thus, it is more likely that the P1 reflected processing of

the target than processing of the preceding gaze cue. Thus, the P1 effect reported by Wykowska et al. may not provide a direct measure of the influence that agency beliefs have on the neural processing of gaze during social interactions.

The occipitotemporal N170, a negative brain potential peaking approximately 170 ms after stimulus presentation, is believed to provide a more sensitive measure of gaze processing than the P1 component. It has been found to be most sensitive to faces and eyes in comparison to inanimate objects (Itier & Taylor, 2004), and it is thought to reflect the earliest structural processing of faces (Ganis et al., 2012). Gaze processing studies have found that the amplitude of the N170 response is influenced by whether gaze is averted or directed at participants, although the direction of this effect is inconsistent (see Itier & Batty, 2009 for review). Interestingly, Pönkänen, Alhoniemi, Leppänen, and Hietanen (2010) found that viewing direct gaze in a live-viewing condition elicited a larger N170 response than averted gaze or closed eyes. This effect was not found when the same faces were viewed as photographs on a computer screen. Pönkänen et al. have argued that this effect suggests that gaze may be processed more "intensely" (p. 486) when it is believed to convey the current perspective, intentions, and agency of another person in real time. However, the outcome of this study provided little insight into the effect that agency beliefs may have on the neural processing of gaze during joint attention interactions.

Carrick, Thompson, Epling, and Puce (2007) addressed this in another ERP study. Participants were presented 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 (the "mutual" condition), or to gaze upwards away from both flanker faces (the "avoid" condition). The onset of the updated central face generated N170 responses measured at occipitotemporal sites. However, these responses were not

modulated by the social significance of the gaze-shift. In contrast, gaze shifts in the group and mutual conditions generated earlier P350 and smaller P500 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 within the depicted social interaction.

The Influence of Agency Beliefs on Gaze Processing During Social Interactions

While previous ERP studies have done well to employ sensitive measures of the neural processing of gaze, they were not designed to investigate the effect of agency beliefs on processing another person's gaze in the context of a social interaction that involves the participant. To this end, we developed a novel virtual reality paradigm to investigate the time course of neural processes associated with evaluating self-initiated joint attention bids (Caruana, de Lissa, et al., 2015). We used this paradigm to measure participants' ERPs while they interacted with a virtual character whom they believed was an avatar controlled by a human in a nearby laboratory via live infrared eye-tracking. Unbeknownst to participants, the virtual character was controlled by a gaze-contingent algorithm. On each trial, participants initiated joint attention towards a task-relevant target. The virtual partner responded by gazing congruently towards the target (achieving joint attention) or incongruently towards one of the remaining on-screen targets (avoiding joint attention). The ERP data revealed that incongruent gaze shifts made by the virtual partner elicited a significantly larger mean centro-parietal P350 ERP than congruent gaze shifts. The same effect was not observed in a non-social control condition that superimposed computer-controlled arrows over the closed eyes of the virtual character. Additionally, this effect was not observed in the N170 data. These data suggest that the P350 ERP is triggered by neural processes associated with the evaluation of socially-related gaze cues. However, it is not clear from this study if the absence of the P350 effect in the control condition was due to the perceptual differences between eyes and arrows, or the fact that

participants did not believe that they were genuinely interacting with another human.

In the current study, we investigated whether the P350 effect identified by Caruana, de Lissa, et al. (2015) was present when participants completed the same task but held an alternative belief that the virtual character was a computer-programmed agent rather than a human-controlled avatar. Data from Caruana, de Lissa, et al.'s (2015) "social" condition became the "avatar" condition in the current study, and we recruited a second group of individuals to participate in an "agent" condition. Consistent with the claim that the P350 ERP represents a process of evaluating the social significance of a gaze shift (e.g., whether joint attention has been achieved), we anticipated that participants in the agent condition would show a significantly reduced P350 effect (i.e., a larger P350 to incongruent gaze shifts than congruent gaze shifts) in the agent condition compared to the avatar condition. We also predicted that smaller occipitotemporal N170 responses would be elicited when participants observed gaze shifts believed to be controlled by a computer agent than a human-operated avatar, irrespective of stimulus congruency. This is consistent with previous studies that have found the gaze-related N170 to be sensitive to human agency beliefs (Pönkänen et al., 2010) but not the social outcome of a gaze shift (Caruana, de Lissa, et al., 2015).

Method

The methods used in this study were approved by the Macquarie University Human Research Ethics Committee.

Participants

This study used an independent-groups design that included two group conditions ("avatar" versus "agent") in which participants responded to two conditions of stimuli ("congruent" versus "incongruent"). Participants volunteered or received course credit for their time and provided written consent before participating.

In the avatar condition, 24 individuals completed the task under the instruction that

211

the virtual character was being controlled by a human partner named "Alan". Participants were instructed that Alan would be interacting with them from a nearby eye-tracking laboratory. The data from two individuals could not be used due to unreliable eye tracking calibration. Another two participants did not believe that the virtual character was being controlled by a human. The behavioural data from a fifth participant also indicated that they were not attending to the virtual character's gaze shifts and were also excluded from analyses (see Behavioural data in Results). This resulted in a final sample of 19 participants (3 male, $M_{age} = 20.95$, $SD = 5.78$) for the avatar condition.

In the agent condition, a separate group of 19 individuals (3 male, $M_{age} = 23.21$, $SD = 6.49$) completed the same task except that they were instructed that the virtual character was a computer-controlled agent. No participants were excluded from the analyses given that reliable behavioural and eye tracking data was obtained for all individuals.

Stimuli

An anthropomorphic virtual character was animated using *FaceGen* (Singular Inversions, 2008). The animated face subtended 8 x 12 degrees of visual angle and was presented in the centre of the screen (a 60 x 34 cm Samsung SynchMaster SA950 HD LED monitor with a refresh rate of 120 Hz) at a distance of 65 cm from the participant. Five face stimuli were generated in which the eyes were either directed at the participant or towards the top-left, top-right, bottom-left or bottom-right corner of the screen. Each corner of the screen contained a cartoon building. These buildings were identical and animated using *GIMP-2* (Kimball & Mattis, 1995). Each building subtended 11 degrees of visual angle. There was 15 degrees of visual angle separating the virtual character's eyes and each building. *Experiment Builder* 1.10.165 (SR Research, 2004) was used to program the gaze-contingent algorithm and present the stimuli.

Stimulus Conditions

We employed the same virtual reality paradigm developed and used in a previous

study (Caruana, de Lissa, et al., 2015). A gaze-contingent algorithm was used to simulate a live interaction between the participant and an on-screen virtual character. Participants believed that the virtual character was controlled by a human partner (avatar condition) or a computer program (agent condition). The tasks completed by participants in the avatar and agent conditions were identical.

Participants were instructed to play a cooperative game with their virtual partner called "Catch the Prisoner". The task was to catch a prisoner who, on each trial, attempted to escape from one of the four prison exits. Participants were told that they would play the role of "watch person" while their virtual partner 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. The watch person was required to inform the guard if a prison exit was breached via initiating joint attention with the guard. A prisoner would be caught if the guard responded congruently to this joint attention bid. However, participants were told that sometimes the guard would respond incongruently to a joint attention bid (and hence the prisoner would escape) because inmates fighting inside the prison distracted him.

At the beginning of each trial, a crosshair was presented in the centre of the screen subtending 1.4 degrees of visual angle. Once the participant (i.e., watch person) fixated for a minimum of 150 ms on the crosshair, it was replaced by an anthropomorphic face of a virtual character (i.e., the guard) with the nasion in the same location as the crosshair. At the same time, four cartoon buildings were displayed in each corner of the screen depicting the prison exits (see Figure 1). After a delay of 200-1000 ms (jittered with a random distribution), a yellow circle (depicting a sensor light that could not be seen by the guard) was presented above one of the exits. The participant was required to look at the spotlight for a minimum of 150 ms. If this was done correctly, after a further delay of 200-1000 ms (jittered with a random distribution), a prisoner appeared at the exit. The participant was then required to initiate joint attention with the guard. To this end, the participant was

required to fixate back on the virtual character's face for at least 150 ms. If this was also done correctly, the guard's gaze shifted after 350-650 ms. This delay provided enough time for an N170 to be generated, but was short enough so that the virtual character's response did not appear unrealistically sluggish. On 50% of trials, the guard's gaze shifted to the correct location (i.e., the escaping prisoner) to achieve joint attention (congruent trials). On the remaining trials, the guard shifted his gaze to one of the remaining three locations (incongruent trials).
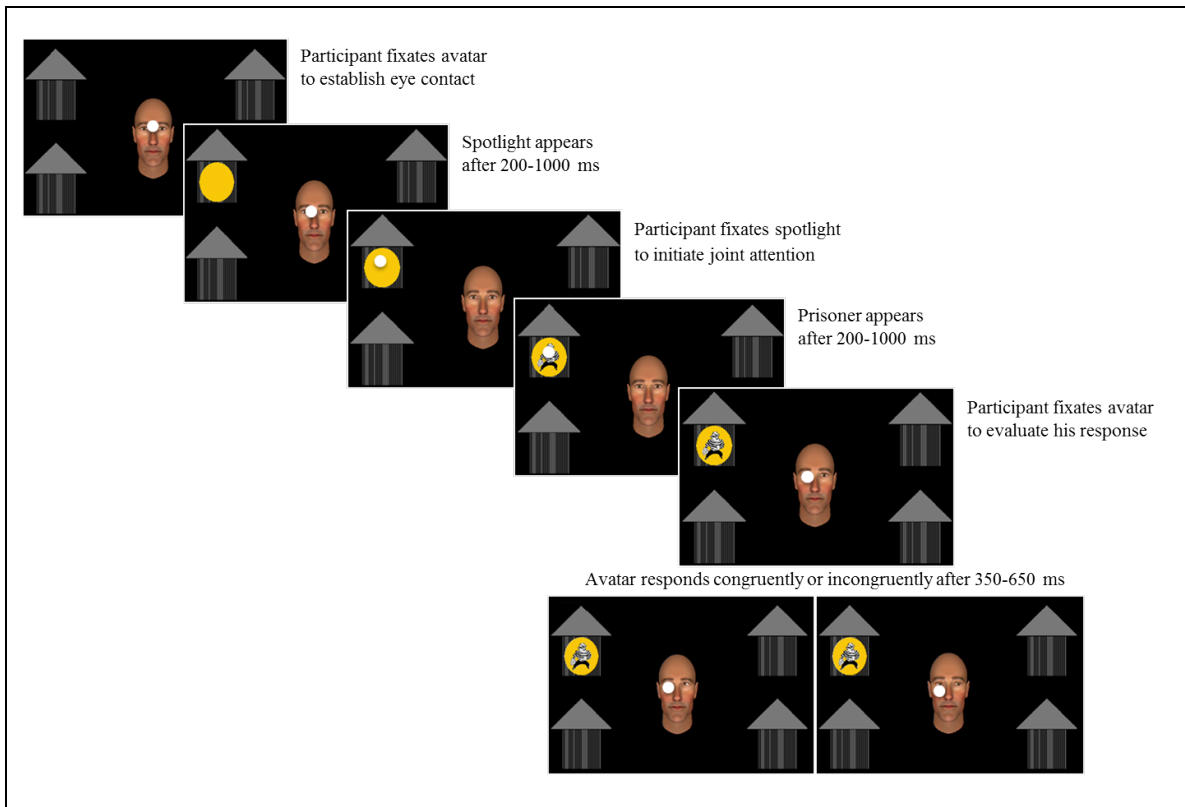
Figure 1. Schematic representation of trial sequence. White circle represents the location of the participant's gaze and was not part of the stimuli visible to the participant.

Participants completed four blocks, each comprising 60 trials. Trials containing congruent and incongruent gaze shifts were presented in random order across blocks. The direction of congruent and incongruent gaze shifts was fully counterbalanced across all trials. Thus, on incongruent trials, the guard was equally likely to gaze towards one of the three buildings not fixated by the participant.

To ensure that participants learned how to engage with the virtual interface appropriately, they received negative feedback (i.e., text reading "Bad Fix" presented in the centre of the screen) if they (1) failed to fixate the spotlight, (2) fixated away from the spotlight before the prisoner appeared, (3) did not fixate back on the guard's face within 3000 ms of the prisoner's appearance, or (4) fixated on the guard's face for less than 1000 ms after fixating the target. This also ensured that participants remained fixated on the guard's face during the interval that gaze-related ERPs were being measured.

Eye Movement and Electroencephalogram (EEG) Recording

An EyeLink 1000 monocular tower-mounted eye tracker was used to record the eye movements of each participant's right eye. Heads were stabilised using a chin rest, and eye movements were sampled at 1000 Hz. The online EEG of each participant was recorded using a Synamps II amplifier with a sampling rate 1000 Hz, an online band pass filter of .05-100 Hz, and a notch filter at 50 Hz. A montage of 29 electrodes were 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. The ground electrode was positioned between FP1, FP2 and FZ. Impedances were maintained below 5 kΩ for all electrodes. Bipolar electrodes were positioned at the outer canthi to measure horizontal electro-ocular activity (HEOG), and above and below the left eye to measure vertical ocular activity (VEOG).

Creating ERPs

Neuroscan 4.5 software was used for the offline processes of the EEG data (Neuroscan, El Paso, Texas, USA). A standard ocular reduction algorithm was used to remove VEOG activity. Corrected data was then band-pass filtered (0.1-30 Hz) with a 12 dB octave roll-off and segmented into epochs that were time-locked to the onset of the virtual character's congruent or incongruent gaze shifts. Epochs comprised a pre-stimulus

baseline of -100 to 0 ms and ended 700 ms after the virtual character's gaze shift. Epochs containing voltages exceeding +/– 100 mV were removed from further analysis. All epochs retained in the analysis were baseline corrected using the 100 ms of pre-stimulus electrical activity. Each participant's accepted epochs were averaged to produce congruent and incongruent ERPs.

Measuring ERPs

To ascertain the length of the intervals used to measure each ERP in this study (P350 at CZ and PZ, and N170 at P7 and P8), we visually inspected the ERPs of each individual. This revealed that a clear P350 peak could be identified in both conditions for 12 individuals from the avatar group and 9 individuals from the agent group. It also revealed clear N170 peaks for all individuals and conditions measured at P7 and P8. Thus, the P350 was measured using mean amplitude whilst the N170 was measured using peak amplitude.

A 130 ms interval (310-440ms) captured each individual's P350 peak in both the congruent and incongruent conditions. Thus, we used 130 ms intervals (65 ms either side of the mean peak) to measure the mean amplitude of P350 (310-440ms) at CZ and PZ and peak amplitude of the N170 (107-237ms) recorded at P7 and P8.

Subjective Experience Questionnaire

At the end of the testing session, participants rated various aspects of their experience on a five-point Likert scale (1 = not at all to 5 = extremely). Participants in both the avatar and agent conditions rated how difficult, intuitive, natural and pleasant the interactive task felt. Following debrief, individuals in the avatar condition were asked to rate how convinced they were that a real person controlled the virtual character. Participants in the agent condition rated the extent to which it felt like they were interacting with a human. They also rated how human-like the virtual character appeared and behaved.

Attention to Gaze Shifts

At the end of each block participants were asked to estimate how frequently (expressed as a percentage) that they successfully caught the prisoner (i.e., the percentage of congruent trials). This provided a measure of task engagement.

Statistical Analysis

To ascertain if parametric or non-parametric analyses should be used, we tested whether (1) datasets within groups for each condition were normally distributed (using the Kolmogorov-Smirnov test), and (2) datasets between groups for each condition were equivariant (using the Levene test). The P350 data passed tests for normality and equivariance. The N170 data failed tests for normality and equivariance. The subjective measures failed tests for normality, and all but one measure passed tests for equivariance.

Since parametric tests are robust to moderate violations of normality when comparing equal samples of this size (Hogg & Tanis, 2010), the effect of group condition (i.e., avatar, agent) on subjective experience ratings were assessed using independent t-tests using statistics that did not assume equal variance. The effect of group condition (i.e., avatar, agent) and stimulus condition (i.e., congruent, incongruent) on each ERP measure was assessed using two-way ANOVAS (Statistical Package for the Social Sciences v19). Main effects of group were assessed using independent t-tests with statistics that did not assume equal variance when the assumption of equivariance was violated. An alpha level of p = 0.05 was used for all analyses.

Results

Subjective Experience Questionnaire

Participants from both the avatar and agent conditions rated the interactive task as easy and intuitive. They also rated the interaction as feeling moderately natural. There were no significant differences between group conditions on these dimensions (independent t-tests, all *p*s > 0.15). However, participants in the avatar condition did rate

the interactive task as significantly less pleasant than participants in the agent condition ($t(26.77) = 2.21, p = .036$). Many participants in the avatar condition explained that they felt frustrated with their partner when he did not respond to their joint attention bid. It is possible that participants in the agent condition did not find this as frustrating given that they knew they were interacting with a computer-programmed agent. The descriptive statistics for these subjective ratings are summarised in Table 1.

Table 1.

Ratings on Subjective Experience Questionnaire

Task Aspect	Avatar	Agent
	M(SD)	*M(SD)*
Difficulty	1.68 (0.75)	1.68 (0.86)
Intuitiveness	4.32 (0.82)	4.32 (0.89)
Naturalness	2.58 (1.26)	3.16 (1.17)
Pleasantness of task*	3.11 (1.66)	4.05 (0.85)

Note. Ratings provided on a 5-point scale (1=low, 5=high). *denotes a significant difference between groups (avatar versus agent).

Avatar condition. All participants in the avatar condition provided ratings that confirmed that they were convinced (5 = completely convinced) that the virtual character was an avatar controlled by a real person ($M = 4.89, SD = 0.32$). Those who provided a 4/5 rating claimed that they only momentarily thought it possible that they were interacting with a computer-controlled agent, and took the interaction for granted.

Agent condition. Participants in the agent condition also rated their level of engagement with the virtual character. On average they reported that the interaction felt moderately human-like ($M = 2.53, SD = 1.02$) and that the virtual character appeared ($M = 3.58, SD = 0.84$) and behaved ($M = 3.47, SD = 0.96$) very human-like.

Attention to Gaze Shifts

At the end of each block, participants estimated the percentage of trials that the virtual character responded congruently. One participant was excluded from the avatar

condition (see Participants in Method) because they provided an average estimate that was two standard deviations above the group mean ($M = 87.00$). The average congruency estimates in the final samples for the avatar ($M = 48.33\%$, $SD = 11.05$) and agent ($M = 48.86\%$, $SD = 77.79$) conditions accurately reflected the 50% congruency manipulation employed in the current study, suggesting that participants were attending to the virtual character's gaze shifts throughout the task.

ERPs

Summary statistics for the mean amplitude measures are shown in Table 2. Group average waveforms comprising the P350 at CZ and PZ are shown in Figure 2, and for the N170 at P7 and P8 are shown in Figure 3. Topographic maps highlighting differences in electrical activity at the scalp between congruent and incongruent condition are depicted for each group condition (i.e., avatar and agent) in Figure 4.

Table 2.

Summary Statistics for Amplitude and Latency Measures by Electrode

	CZ		PZ	
	Congruent	**Incongruent**	**Congruent**	**Incongruent**
P350 mean amplitude				
Avatar	11.85(4.60)	14.90(5.45)	10.52(4.15)	12.53(4.53)
Agent	8.44(4.52)	8.69(5.23)	6.43(3.48)	6.46(4.31)
	P7		**P8**	
	Congruent	**Incongruent**	**Congruent**	**Incongruent**
N170 peak amplitude				
Avatar	-7.42(4.59)	-7.24(4.27)	-9.79(5.48)	-9.89(5.05)
Agent	-4.72(2.27)	-4.93(2.00)	-7.53(5.74)	-7.47(5.55)

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

P350 mean amplitude. There was a main effect of group measured at both CZ ($F(1,36) = 9.492$, $p = .004$), and PZ ($F(1,36) = 15.492$, $p < .005$) since the P350 generated in the avatar condition was significantly larger than the agent condition. There was also a main effect of condition measured at CZ ($F(1,36) = 17.605$, $p < .005$), and PZ ($F(1,36) = 7.790$, $p = .008$) because the P350 was significantly larger on incongruent trials than congruent trials. Most importantly, there was a significant interaction between group and condition at CZ ($F(1,36) = 12.739$, $p = .001$) and PZ ($F(1,36) = 7.272$, $p = .001$) because the difference between the P350 in the incongruent and congruent conditions was larger in the avatar condition [CZ: ($t(18) = 4.798$, $p < .001$; PZ: ($t(18) = 3.425$, $p = .003$] than the agent condition [(CZ: ($t(18) = .533$, $p = .600$); PZ: ($t(18) = .079$, $p = .938$].

N170 peak amplitude. There was a main effect of group at P7 ($F(1,36) = 5.10$, $p = .030$) since the N170 generated in the avatar condition was significantly larger than the agent condition. This was also significant when assessed using an independent t-test not assuming equal variance ($t(25.63) = 2.258$, $p = .033$). A main effect of group was not significant at P8 ($F(1,36) = 1.763$, $p = .193$). Similarly, there was no significant main effect of condition [(P7: ($F(1,36) = .006$, $p = .937$); P8: ($F(1,36) = .008$, $p = .931$)], and no significant group*condition interaction when measured at either P7 ($F(1,36) = 0.892$, $p = .351$) or P8 ($F(1,36) = 0.160$, $p = .692$).
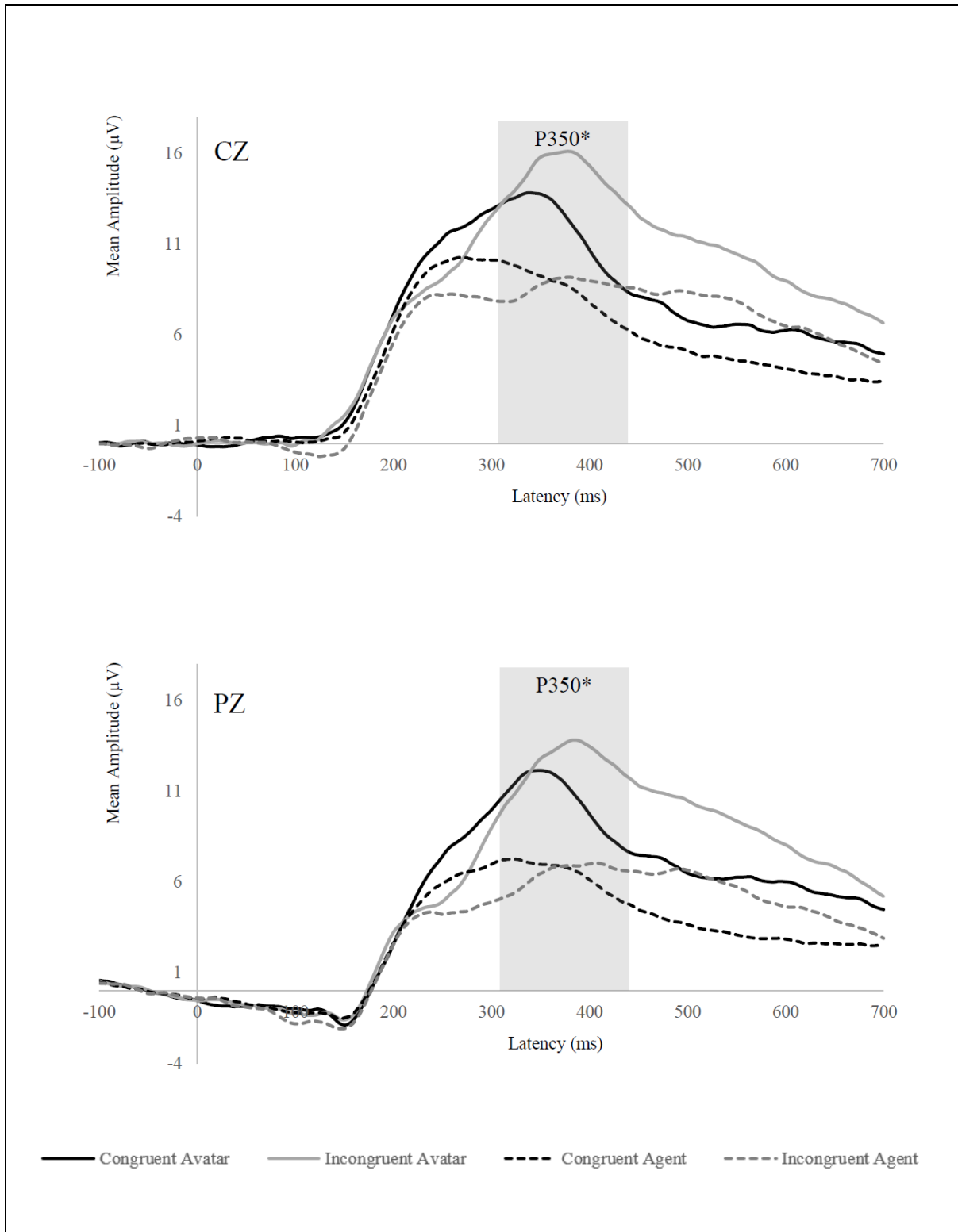
Figure 2. Group average waveforms comprising the P350 at (a) Cz and (b) Pz electrodes. Epochs were time-locked to the onset of the virtual character's gaze shift.
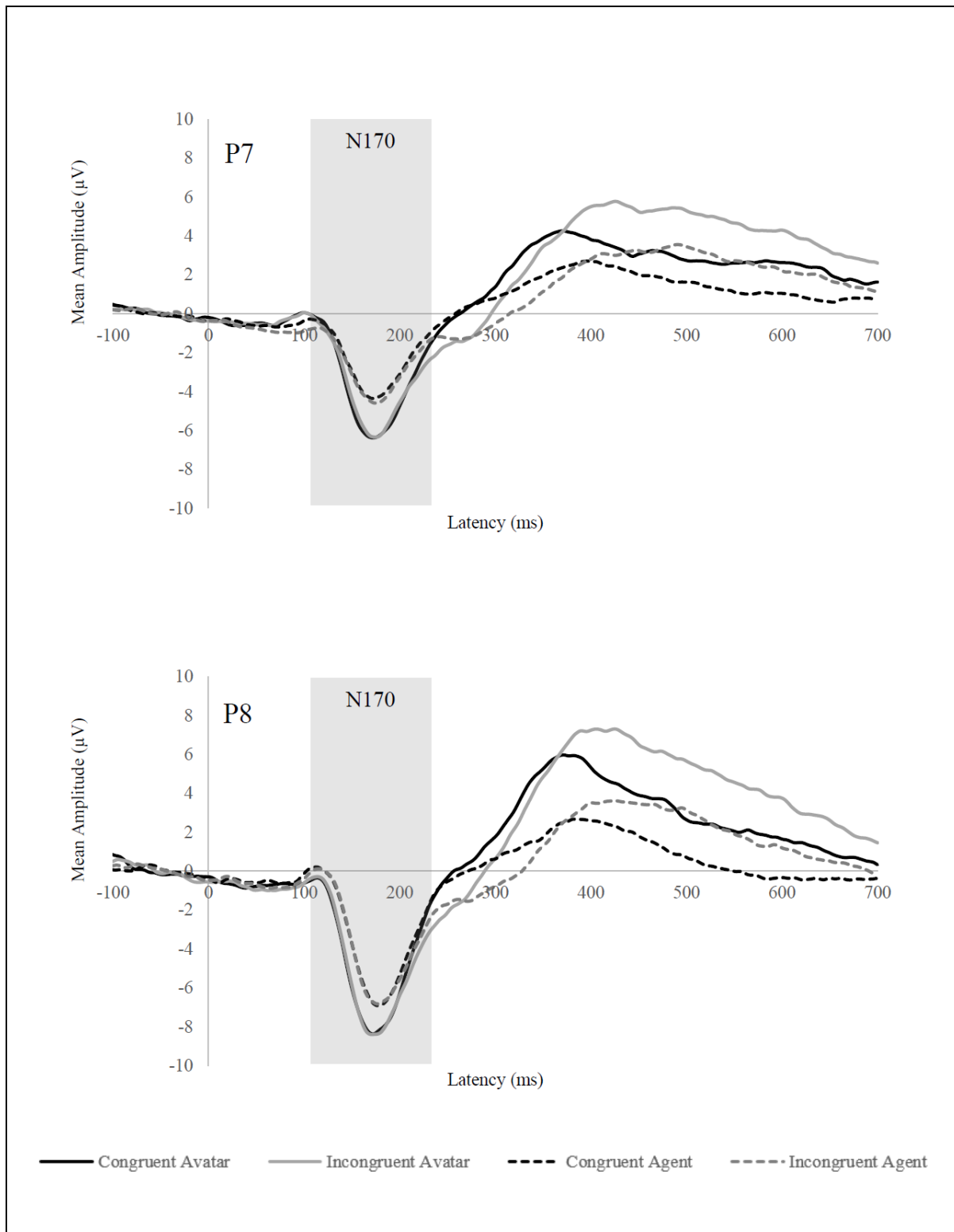
Figure 3. Group average waveforms comprising the N170 at (a) P7 and (b) P8 electrodes. Epochs were time-locked to the onset of the virtual character's gaze shift.
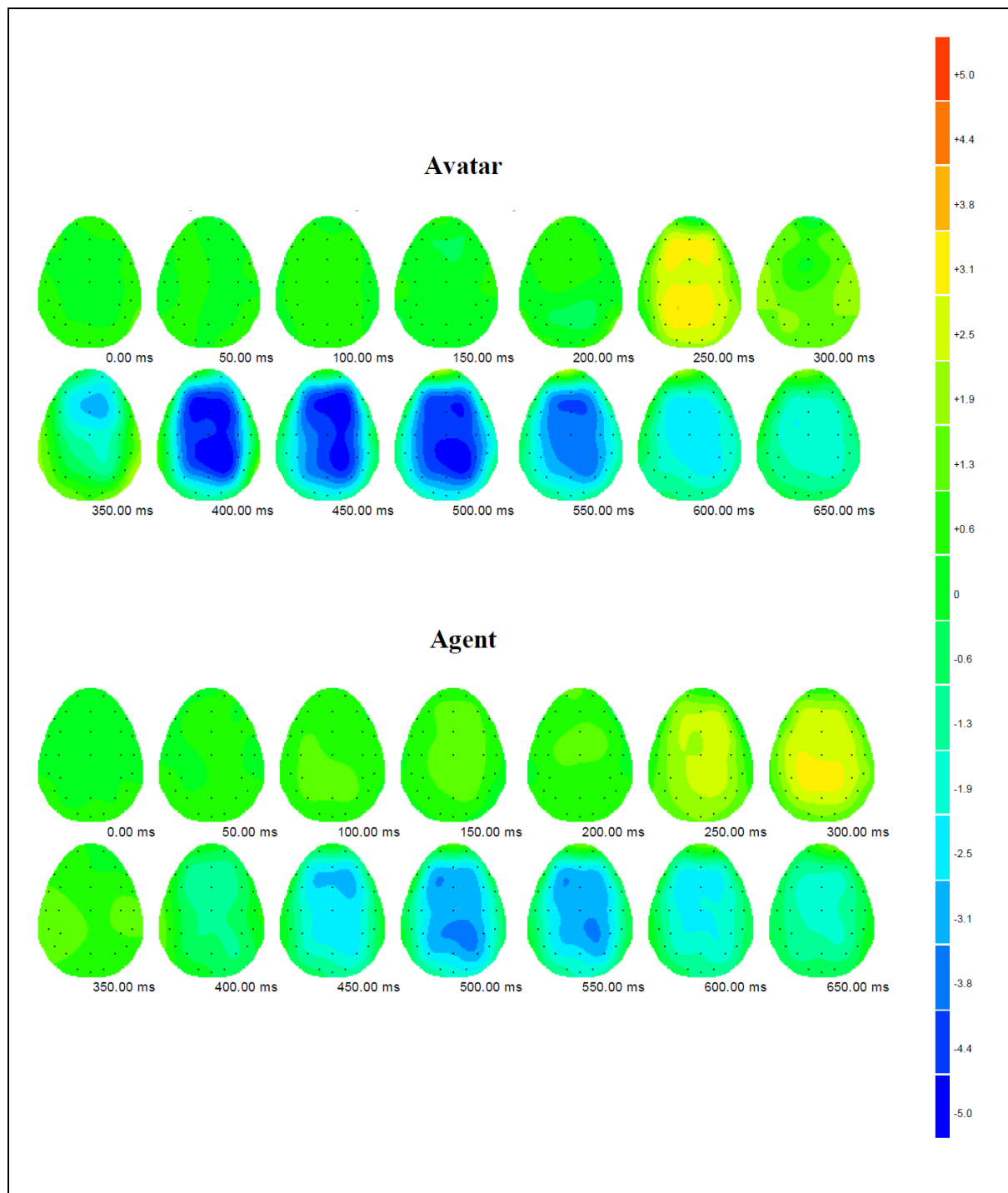
Figure 4. Effect topography maps (Congruent – Incongruent) by group.

Discussion

The aim of the current study was to determine if agency beliefs influence the neural processes associated with evaluating the achievement of joint attention during gaze-based social interactions. We predicted that the centro-parietal P350 effect, previously identified by Caruana, de Lissa, et al. (2015), would be significantly larger in participants who believed that they were interacting with a human than those who believed that they were interacting with a computer. We also predicted that a significantly larger N170 would be evoked by gaze shifts believed to be made by a human than a computer, regardless of whether the gaze shift resulted in joint attention or not. Consistent with these predictions, the centro-parietal P350 ERP only differed between congruent and incongruent gaze shifts when individuals believed the virtual character to be controlled by a human avatar. In addition, larger occipitotemporal N170 responses were observed in individuals who believed that they were interacting with a human than individuals who believed that they were interacting with a computer. Taken together, these data suggest that agency beliefs influence the neural processing of social signals conveyed by virtual characters.

The Influence of Agency Beliefs on Processing the Social Outcome of Gaze Shifts

The P350 ERP effect measured at centro-parietal electrodes was only observed in the avatar condition in participants who believed that the virtual character's gaze was controlled by a human. This centro-parietal P350 effect has been previously associated with evaluating the social significance of a gaze shift (Carrick et al., 2007). In the current study, it was specifically associated with evaluating whether a gaze shift signalled the achievement of joint attention with another person (i.e., "Is my partner attending the same thing as me?"). If the P350 ERP truly represents the onset of mentalising during gaze processing, it makes sense that this effect was not present when the participant did not believe the gaze shift to represent the intentional actions of another human. In line with this expectation, the P350 effect was absent in participants who believed that they were

224

interacting with a computer-programmed agent.

This finding provides support for the social-specificity of the P350 effect that we identified in our earlier work (Caruana, de Lissa, et al., 2015). Given that the P350 effect was only observed in the avatar condition, it is unlikely that it represents an effect of gaze congruency on non-social attention mechanisms (e.g., odd-ball, error detection, or attention orienting effects). If this were the case, the P350 effect should have been measured in both the avatar and agent conditions, since both conditions manipulate the spatial properties of gaze in the same way. Thus the P350 effect seems specific to conditions where participants believe the virtual character is a human-controlled avatar. This may be because the P350 ERP reflects the evaluation of gaze to represent another person's mental perspective.

The Influence of Agency Beliefs on the Perceptual Processing of Gaze Shifts

Consistent with previous findings, larger N170 responses to gaze shifts were measured in individuals who believed that they were interacting with a human rather than a computer (Pönkänen et al., 2010). Wykowska et al. (2014) have argued that these effects of agency beliefs on the early perceptual processing of gaze-related stimuli may be driven by neural mechanisms of "stimulus gain control". Specifically, neural processing of sensory information may be amplified to increase the signal-to-noise ratio for stimuli that are relevant to the observer's current context. This has been explained using the Intentional Stance Model proposed by Wykowska et al. This model argues that the brain takes an "intentional stance" towards stimuli believed to represent a human mind. This involves the recruitment of neural substrates that govern mentalising processes (e.g., medial prefrontal cortex, and temporoparietal junction). These mentalising mechanisms may then have a top-down influence on attentional control in the parietal cortex (e.g., intraparietal sulcus) by prioritising the processing of social stimuli. This results in the enhanced early processing of social stimuli in extrastriate visual areas where the sensory gain effect is measured in occipitotemporal ERPs (e.g., P1 and N170). In the current study, the N170 effect only

reached significance in the left hemisphere. This is consistent with lesion studies that have reported that the left temporoparietal junction is especially important in supporting the ability to evaluate another's mental perspective (Samson et al., 2004). There is also evidence that individual differences in anthropomorphism (i.e., the tendency to attribute human agency to non-human phenomenon) are correlated with increased grey matter volume in the left temporoparietal junction (Cullen, Kanai, Bahrami, & Rees, 2013).

Whilst the Intentional Stance Model provides a sensible framework for interpreting these converging findings, the evidence supporting the direction of the proposed top-down relationship between mentalising and early visual perception brain regions remains tentative. Future research integrating neuroimaging techniques that have high temporal (e.g., EEG, MEG) and spatial (e.g., fMRI, PET) resolution are needed to determine whether the Intentional Stance Model provides an accurate account of the mechanisms underlying the influence of agency beliefs on early perceptual processes. Specifically, connectivity analyses and dynamic causal modelling may elucidate whether neural substrates associated with mentalising modulate occipitotemporal and parietal areas early on in the perceptual processing of gaze shifts. Whilst the current study cannot confirm all of the mechanisms proposed by the Intentional Stance Model, it does corroborate the finding that gaze-related N170 responses are modulated by beliefs of human agency (Pönkänen et al., 2010).

Implications and Recommendations

These findings present both methodological and empirical implications for social neuroscience research. First, this study contributes further evidence that beliefs about human agency influence the neural processing of social signals conveyed by virtual characters. This suggests that in order to achieve an ecologically valid simulation of social interactions using virtual characters, participants must believe the virtual character to represent a real human whom they can attribute mental states to. Not only does this match

our subjective experience during real social interactions, but this belief is important in engaging the neural processes that support genuine social interactions (e.g., mentalising). Therefore, our data suggests that the practical and ethical considerations involved in deceiving participants are justified by the importance of this benign deception in supporting the ecological validity of virtual interactions.

Second, the current study provides a new approach for achieving control over the effects that gaze stimuli may have on non-social cognitive processes. Traditionally, gaze-processing studies have relied on arrow stimuli to control for the effects that gaze may have on spatial attention (see Nation & Penny, 2008 for review). However, it is impossible to obtain gaze and arrow stimuli that are perceptually equivalent. This is reflected by the inconsistent cueing effects found in paradigms comparing gaze and arrow cues in behavioural (see Frischen, Bayliss, & Tipper, 2007 for review) and ERP studies (e.g., Feng & Zhang, 2014; Hietanen, Nummenmaa, Nyman, Parkkola, & Hämäläinen, 2006; Holmes, Mogg, Garcia, & Bradley, 2010; Lassalle & Itier, 2013; van Velzen & Eimer, 2003). We suggest that the social-specificity of gaze-related ERP effects can be conservatively determined by manipulating whether the gaze stimuli is believed to be controlled by a human or computer, rather than manipulating whether spatial information is conveyed by eyes or arrows.

In addition to achieving an ecologically valid measure of the neural processing of gaze shifts, inducing agency beliefs may assist in minimising the effect that individual differences in anthropomorphism have on gaze-related ERPs. For example, it is possible that the current study observed a P350 effect in the avatar condition but not the agent condition because an agency belief ensures that participants consistently treat the virtual character as a human. Given that individuals differ in their propensity to anthropomorphise non-human stimuli (e.g., Cullen et al., 2013), it is possible that the P350 effect may have been present to different extents in some individuals in the agent condition. That is, some

individuals may have been more likely to anthropomorphise the virtual character, resulting in the engagement of spontaneous mentalising processes. This is consistent with previous findings that individual differences in anthropomorphism are correlated with the size of brain regions associated with mentalising (Cullen et al.). Whilst future virtual reality studies could employ measures of anthropomorphism as a covariate to account for these individual differences, this source of noise can be effectively minimised by ensuring participants believe the virtual character to be a human-controlled avatar.

Summary

In sum, the current study demonstrates that the neural processing of gaze is sensitive to agency beliefs. This has significant implications for the use of virtual reality as a tool for simulating ecologically valid interactions in social neuroscience research. We found larger left occipitotemporal N170 responses to gaze shifts in individuals who believed the virtual character's gaze shift to be controlled by a human rather than a computer. This suggests that agency beliefs may have a top-down influence on the early perceptual processing of gaze. Furthermore, we found that a centro-parietal peak differentiated gaze shifts that signalled the success or failure of a self-initiated joint attention bid after approximately 350 ms. This P350 effect was only observed in individuals who believed the virtual character to be operated by a human. These data support the claim that the brain begins to decode information about whether a human's focus of attention is the same or different to our own approximately 350 ms after the observation of a gaze shift. Thus the P350 may provide a useful neural marker for evaluating the achievement of joint attention, which may be used in future research investigating how gaze is processed by individuals with autism. It would be specifically interesting to investigate the relationship between the P350 and social communication ability, and whether clinical gains in social communication intervention programs are associated with changes in the P350 effect.

References

Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. A. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of Autism & Developmental Disorders, 39*(1), 84-96. doi: 10.1007/s10803-008-0601-7

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

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. doi: 10.1097/WNR.0b013e3280ebb44b

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

Caruana, N., de Lissa, P., & McArthur, G. (2015). The neural time course of evaluating self-initiated joint attention bids. *Brain and Cognition*. doi: http://dx.doi.org/10.1016/j.bandc.2015.06.001

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions Royal Society London Biological Sciences, 358*, 315-324. doi: 10.1098/rstb.2002.1199

Cullen, H., Kanai, R., Bahrami, B., & Rees, G. (2013). Individual differences in anthropomorphic attributions and human brain structure. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nst109

Feng, Q., & Zhang, X. (2014). Eye gaze triggers reflexive attention shifts: evidence from lateralised ERPs. *Brain Research, 1589*, 37-44. doi: 10.1016/j.brainres.2014.09.029

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. doi: 10.1037/0033-2909.133.4.694

Ganis, G., et al. (2012). The N170, not the P1, indexes the earliest time for categorical perception of faces, regardless of interstimulus variance. *NeuroImage, 62*(3): 1563-1574. doi: 10.1016/j.neuroimage.2012.05.043

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., & Vogeley, K. (2014). The use of virtual characters to assess and train nonverbal communication in high-functioning autism. *Frontiers in Human Neuroscience, 8.* doi: http://dx.doi.org/10.3389/fnhum.2014.00807

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. doi: 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. doi: 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. doi: http://dx.doi.org/10.1016/j.neubiorev.2009.02.004

Itier, R. J., & Taylor, M. J. (2004). Face recognition memory and configural processing: a developmental ERP study using upright, inverted, and contrast-reversed faces. *Journal of Cognitive Neuroscience, 16*(3), 487-502. doi: 10.1162/089892904322926818

Kimball, S., & Mattis, P. (1995). GIMP Image Manipulation Program (Version 2.8.2). Berkley.

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. doi: 10.1080/17470919.2013.835750

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research, 2*(1), 2-21. doi: 10.1002/aur.61

Murray, D. S., Creaghead, N. A., Manning-Courtney, P., Shear, P. K., Bean, J., & Prendeville, J. A. (2008). The relationship between joint attention and language in children with autism spectrum disorders. *Focus on Autism & Other Developmental Disabilities, 23*(1), 5-14. doi: 10.1177/1088357607311443

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? Is it automatic? Is it social? *Development and Psychopathology, 20*(01), 79-97. doi: doi:10.1017/S0954579408000047

Pelphrey, K. A., Shultz, S., Hudac, C. M., & Vander Wyk, B. C. (2011). Research review: Constraining heterogeneity: the social brain and its development in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 52*(6), 631-644. doi: 10.1111/j.1469-7610.2010.02349.x

Peter, J., Valkenburg, P. M., & Schouten, A. P. (2005). Developing a model of adolescent friendship formation on the internet. *Cyberpsychological Behaviour, 8*(5), 423-430. doi: 10.1089/cpb.2005.8.423

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., & Vogeley, K. (2014). Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *NeuroImage, 101*, 124-137. doi: 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.* doi: 10.1093/scan/nsq068

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*(4), 393-414. doi: 10.1017/S0140525X12000660.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah N. J., Fink, G. R., & Vogeley, K. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience, 22*(12), 2702-2715. doi: 10.1162/jocn.2009.21401.

Singular Inversions. (2008). FaceGen Modeller (Version 3.3) [Computer Software]. Toronto, ON: Singular Inversions.

SR Research. (2004). Experiment Builder (Version 1.10.165). Ontario.

Tomasello, M. (1995). Joint Attention as Social Cognition. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development*. Hillsdale: Lawrence Erlbaum Associates

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. doi: 10.1111/1469-8986.00083

von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S.-H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior, 26*(6), 1641-1650. doi: http://dx.doi.org/10.1016/j.chb.2010.06.012

Watson, D., Tellegen, A., & Clark, L. A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scale. Journal of Personality and Social Psychology, 54, 1063-1070. doi: http://dx.doi.org/10.1037/0022-3514.54.6.1063

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. doi: 10.1371/journal.pone.0094339

Chapter 7

General Discussion

General Discussion

Introduction

Joint attention is a pivotal cognitive ability that is a precursor to the development of language and social cognition, and is a cornerstone in our everyday social interactions with others (Adamson, Bakeman, Deckner, & Romski, 2009; Baron-Cohen, 1995; Charman, 2003; Mundy, Sigman, & Kasari, 1990; Murray et al., 2008). Impairments in joint attention are a characteristic of autism spectrum disorders, and are related to outcomes in language development and social functioning (Charman, 2003; Dawson et al., 2004; Mundy et al., 1990).

Whilst much is known about the development of joint attention, there has been little empirical investigation of the cognitive and neural mechanisms underlying joint attention in typical development and in autism. This is largely due to a lack of ecologically valid experimental paradigms that measure joint attention. To address this, Schilbach and colleagues (2013) proposed a "second person" approach to social cognition and neuroscience research, whereby the primary objective is to develop paradigms that provide experimentally controlled and objective measures of social cognition during ecologically valid social interactions. The second person approach is particularly pertinent to the measurement of joint attention because joint attention can only be experienced in the context of a social interaction.

Inspired by this second person approach, I have addressed two overarching research aims in my dissertation. My first aim was to develop a second person paradigm for the measurement of joint attention in behavioural and neuroimaging contexts, including the ability to respond to joint attention bids (RJA), initiate joint attention bids (IJA) and evaluate the achievement of joint attention bids (EAJA). To this end, in **Chapter 2**, I reviewed the existing experimental approaches for measuring joint attention, and derived the critical features necessary for a second person measurement of joint attention. I also

presented a new "Catch-the-Burglar" virtual reality paradigm that integrated these features to measure the behaviours and neural processes associated with joint attention.

The second aim of my dissertation was to apply the Catch-the-Burglar paradigm to investigate the cognitive and neural mechanisms of joint attention in people with typical development and in individuals with autism. In **Chapter 3**, I applied my paradigm in a functional magnetic resonance imaging (fMRI) study to investigate the neural correlates of RJA and IJA in adults with typical development. In **Chapter 4**, I extended the use of this paradigm in an eye tracking study to compare objective measures of RJA and IJA performance in typical development and in autism. In **Chapter 5**, I modified my paradigm and applied it in an event-related potential (ERP) study which investigated the neural time course of EAJA. In **Chapter 6**, I used the same ERP paradigm to examine whether beliefs about a virtual character's agency (i.e., human-controlled or computer-controlled) influenced gaze-related ERPs during joint attention interactions.

A New Second Person Approach to the Measurement of Joint Attention

In **Chapter 2**, I reviewed the existing approaches to measuring joint attention in behavioural and neuroimaging contexts, and discussed the importance of taking a second person approach to social cognition research (cf. Schilbach et al., 2013). I highlighted the innovations of recent studies that have developed interactive joint attention paradigms (e.g., Redcay et al., 2012; Redcay et al., 2010; Schilbach et al., 2010). My aim was to integrate these innovations within a new paradigm to achieve a better balance between ecological validity and experimental control in the measurement of joint attention. I was particularly inspired by the idea of simulating cooperative interactions in the laboratory, which motivate intentional and goal-directed joint attention behaviours (cf. Redcay et al., 2012; Redcay et al., 2010). My paradigm was also inspired by the previous use of gaze-contingent virtual interfaces to simulate seemingly genuine yet experimentally controlled interactions (cf. Schilbach et al., 2010; Wilms et al., 2010).

As well as merging the strengths of previous joint attention paradigms, the Catch-the-Burglar paradigm incorporates a number of new innovations to further enhance the ecological validity of experimental joint attention interactions. The most unique feature of my paradigm is the "search phase" at the beginning of each trial in which the participant (and their virtual partner) are required to search the virtual environment for the target of joint attention (i.e., the burglar). This is important for two reasons. First, it provides a context in which participants can implicitly determine their social role – as an "initiator" or "responder" in the joint attention episode – throughout the course of the trial, depending on whether or not they find a target. This contrasts with previous paradigms in which participants have been explicitly instructed about their social role before the trial began (e.g., Redcay et al., 2012; Redcay et al., 2010; Schilbach et al., 2010). Second, the search phase in my paradigm establishes a context in which the virtual character displays non-communicative "searching" gaze shifts. Thus, participants are required to establish and interpret eye contact as an ostensive cue to disambiguate communicative and non-communicative gaze shifts. As in everyday social interactions, these processes of attention and intention monitoring allow individuals to appropriately identify opportunities for joint attention. It is of clinical importance that both of these processes are included in the measurement of joint attention given that their failure may give rise to joint attention impairments in autism. Thus, compared to previous approaches, my paradigm provides a more ecologically valid measure of joint attention behaviours by capturing the processes of (1) determining one's role as a responder or initiator of a joint attention bid, and (2) identifying appropriate opportunities for joint attention by monitoring the attention and intentions of one's social partner.

In addition to providing a more ecologically valid measure of joint attention, I designed closely matched baseline conditions to provide greater experimental control over the non-social task demands that may influence behaviour and neural processing. This is

particularly important for the application of joint attention paradigms in fMRI studies that depend upon the subtraction method to identify the neural correlates of joint attention. The subtraction method relies on the assumption of pure insertion, which is the idea that two conditions (i.e., test and baseline conditions) only differ with respect to the cognitive process of interest. However, as demonstrated in the critical review provided in **Chapter 2**, previous joint attention studies have not met this assumption, making it difficult to interpret whether the neural correlates they identify specifically represent the neural and cognitive mechanisms of joint attention. I believe that the baseline conditions employed in the Catch-the-Burglar paradigm come closer to satisfying the principle of pure insertion. Thus, this paradigm could be used to address a number of research questions about the cognitive and neural mechanisms of joint attention, such as, *What are the neural correlates specific and common to RJA and IJA?*, and *Are joint attention difficulties in autism independent of impairments in attention or oculomotor control?* I addressed these research questions in studies outlined in **Chapter 3** and **Chapter 4**, respectively.

Main Experimental Findings

Responding to and Initiating Joint Attention Bids (RJA and IJA)

In **Chapter 3**, I used the Catch-the-Burglar paradigm in an fMRI study that investigated the neural correlates associated with RJA and IJA. Both RJA and IJA behaviours were associated with a broad pattern of activation across a frontotemporoparietal network, consistent with the social brain network identified in previous neuroimaging studies of joint attention (see Pfeiffer, Vogeley, & Schilbach, 2013 for review). The closely matched baseline conditions allowed me to directly compare activation associated with RJA and IJA. A conjunction analysis revealed that a right-lateralised subset of the identified frontotemporoparietal network was commonly involved when participants engaged in RJA and IJA behaviours. This included 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.

This was the first study to show direct evidence in support of the Parallel Distributed Processing Model (PDPM; Mundy, Sullivan, & Mastergeorge, 2009). This model postulates that RJA and IJA are predominantly supported by posterior-parietal and anterior networks, respectively. These networks are argued to integrate throughout development to support the parallel processing of self- and other-oriented representations during social interactions. This may include representing the locus of agency during the interaction (e.g., "I am initiating and my partner is responding") or comparing visual perspectives (e.g., "My partner is looking at an object that I cannot see"). My data supports this claim since (1) both RJA and IJA trials were found to activate common substrates across a frontotemporoparietal network, and (2) these substrates – specifically MFG, IFG, TPJ and precuneus – have been previously implicated in tasks that involve the simultaneous representation of self- and other-oriented representations (e.g., Halko, Hlushchuk, Hari, & Schürmann, 2009; Kincade, Abrams, Astafiev, Shulman, & Corbetta, 2005; Samson, Apperly, Chiavarino, & Humphreys, 2004; Saxe, Schulz, & Jiang, 2006; Vogeley et al., 2004). The ability to simultaneously represent self- and other-oriented perspectives is important to achieve joint attention during a dynamic social interaction. This is because we must continually adapt our behaviour to align with our social partner's changing attentional perspective, intentions, and actions so that we can coordinate our behaviours and achieve a common focus of attention.

In the same study, I also found that IJA trials triggered greater activation of the anterior cingulate cortex (ACC) than RJA trials. The ACC has been previously associated with executing goal-directed behaviours and action monitoring (Shackman et al., 2011). Thus, it is possible that the activation of the ACC during IJA trials in the Catch-the-Burglar paradigm reflects the attention and intention monitoring processes involved when

evaluating whether a social partner is ready to follow a joint attention bid (i.e., by establishing eye contact). This is the first study to implicate the ACC as a neural correlate of joint attention (cf. Redcay et al., 2012; Schilbach et al., 2010). This may be because the Catch-the-Burglar paradigm provides a more ecologically valid measure of joint attention than previous studies by capturing the attention and intention monitoring processes that pre-empt adaptive RJA and IJA behaviours.

In **Chapter 4**, I report on a study that used the Catch-the-Burglar paradigm to compare RJA and IJA behaviours in adults with high-functioning autism (HFA) to adults with typical development (controls). I also tested whether any difference between individuals with HFA and controls were specifically related to the social cognitive processes involved in achieving joint attention, or whether they reflected non-social cognitive deficits (e.g., oculomotor control, action inhibition, and attention orienting). I achieved this by comparing performance on RJA and IJA trials to baseline trials which had identical non-social task demands. I found that adults with HFA made significantly more errors when responding to and initiating joint attention bids compared to controls. Interestingly, impairments in RJA were observed in social RJA trials but not baseline trials. This suggested that the RJA impairments observed in participants with HFA represented an impairment in social cognition rather than an impairment in non-social cognition.

I also found that adults with HFA were slower than controls to execute saccades to respond to joint attention bids, even on trials where they succeeded in achieving joint attention. Again, this effect was specifically observed on RJA trials but not baseline trials. Interestingly, this group difference diminished by the end of the experiment because most individuals with HFA were able to adapt to the task, and their response times decreased to match those of controls in the final block of the experiment.

In contrast to the effects observed for RJA, I found that the group differences in

IJA were not specific to the social condition. Rather, individuals with HFA were significantly poorer than controls in both the social and baseline IJA conditions. However, there was a high level of individual variation in performance within the HFA group. This was particularly evident in the analysis of IJA behaviours on trials where participants succeeded in initiating joint attention. Eye movement analyses revealed that some individuals with HFA were exceptionally slow to establish eye contact with their partner before attempting to initiate joint attention. This was characterised by longer fixation times on the target of joint attention (i.e., the burglar) and an increased tendency to initiate joint attention before establishing eye contact. For many individuals with HFA, these behaviours were only observed in the social IJA condition. It is interesting to note that the impaired IJA performance in some individuals with autism was quite severe. Specifically, some individuals with autism fixated the target of joint attention between 2-6 seconds on average before attempting to initiate joint attention, compared to less than 1 second on average in controls. Taken together, the findings from **Chapter 3** and **Chapter 4** demonstrate that the Catch-the-Burglar paradigm is a valid and sensitive measure of the cognitive and neural mechanisms of RJA and IJA in people with typical development and people with HFA.

Evaluating the Achievement of Joint Attention (EAJA)

During social interactions, we constantly evaluate the significance of our social partner's gaze shifts in order to coordinate our behaviour with their current perspective. In a study discussed in **Chapter 5**, I used the Catch-the-Burglar paradigm to conduct the first investigation of the time course of neural processes associated with EAJA (Caruana, de Lissa, & McArthur, 2015). The ability to evaluate the achievement of joint attention is important in the coordination of joint attention during dynamic social interactions. This is because EAJA enables individuals to identify whether they have succeeded in sharing information with another person, or whether another communicative attempt is required. In

this study, participants engaged in a modified version of the Catch-the-Burglar paradigm. Participants were asked to play a cooperative game with a virtual partner that required them to initiate joint attention towards a target on each trial. In response, the virtual partner shifted his gaze either congruently towards the target (achieving joint attention) or incongruently towards a different location (avoiding joint attention). Event-related potentials (ERPs) that were time-locked to the onset of the virtual partner's gaze revealed a positive voltage increase over centro-parietal cortical sites peaking after approximately 350 ms. This P350 peak was significantly larger and later when the partner's gaze shift resulted in the avoidance of joint attention than the achievement of joint attention. Further, it was completely absent in a second group of participants who completed a non-social analogue of the same task, in which an arrow stimulus replaced the virtual partner's gaze response. This suggested that the P350 ERP reflected an effect of social cognitive processes, and not a non-social effect related to attention modulation or error detection.

Given that the P350 effect identified in **Chapter 5** was not observed in the control condition, where arrow stimuli replaced the virtual character's gaze, this ERP could be interpreted as specifically reflecting the time course of EAJA processes. However, the strength of this interpretation is mitigated by the fact that arrows and gaze are not perceptually equivalent stimuli. Thus, in **Chapter 6**, I compared the ERPs of participants from the study in **Chapter 5** – who were told that the virtual character represented a real person (i.e., avatar condition) – to the ERPs of a new group of participants who were informed that the virtual character was controlled by a computer program (i.e., agent condition). Interestingly, not only was the P350 significantly smaller in the latter group than in the former group, but the left occipitotemporal N170 responses elicited by the observation of the virtual character's gaze shift were significantly larger (more negative) when the virtual character was believed to be a human-controlled avatar rather than a computer-programmed agent. These findings are consistent with the idea that perceiving a

genuine interaction may recruit neural substrates that support mentalising (i.e., the ability to understand the mental states of others), which may have a top-down influence on how social information is perceived and evaluated (Wykowska, Wiese, Prosser, & Müller, 2014). These findings also suggest that the P350 may provide a neural marker that could be used in future studies as a biomarker for gaze processing deficits in individuals with autism, or as an objective outcome measure for interventions targeting social communication skills. Additionally, the outcomes of the study presented in **Chapter 6** confirm that the P350 effect is specific to contexts in which participants believe that they are engaged in a genuine social interaction. This finding provides a theoretical contribution by elucidating the influence that agency beliefs have on the P350, and offers important methodological insights into the application of second person paradigms in a way that is both ethical and ecologically-valid.

Future Directions

The Measurement of Joint Attention

The Catch-the-Burglar paradigm presented in **Chapter 2**, and applied in **Chapters 3**, **4**, **5** and **6**, has made several advances towards the measurement of RJA, IJA and EAJA in experimental settings. However, more can be done to achieve greater ecological validity in the measurement of joint attention. For instance, whilst the Catch-the-Burglar paradigm is the first to provide an objective and sensitive measure of IJA behaviours using eye tracking, there are outstanding questions regarding the most appropriate way to define and measure IJA performance. Interestingly, I found in two separate studies that adults with typical development often made a saccade to initiate joint attention before establishing eye contact with their partner (Chapter 3 & 4; Caruana, Brock, & Woolgar, 2015). I also found that this tendency to prematurely initiate joint attention was exacerbated in a subgroup of individuals with HFA (Chapter 4). In the baseline condition, by contrast, participants almost always waited for the fixation point to turn green before saccading towards the

burglar's location. This difference between conditions may have occurred because participants expected their "human" partner to follow their gaze even though eye contact had not been established.

Given that the Catch-the-Burglar paradigm is the first to employ an eye-tracking measure of IJA, further work is needed to validate the way in which we have defined and measured adaptive IJA performance. Specifically, we have considered the establishment of eye contact before IJA as a necessary step in successfully demonstrating IJA behaviour. Our approach to measuring IJA could be assessed in studies comparing IJA behaviours measured using our paradigm with observations of IJA behaviour between adults in naturalistic face-to-face interactions. Determining the most appropriate definition and measure of IJA performance is tightly bound to conceptual questions regarding the definition of joint attention. For instance – *Can joint attention be initiated by staring at an object, instead of establishing eye contact first?* and *When is it appropriate or maladaptive to initiate joint attention without establishing eye contact first?* Answers to these questions will help guide the future development of gaze-contingent paradigms that can provide ecologically-valid measures of IJA behaviour. For instance, future studies could develop more flexible gaze-contingent algorithms in which virtual characters are programmed to respond to a participant's joint attention bid regardless of whether eye contact is established.

The ecological validity of joint attention measures, like the Catch-the-Burglar paradigm, could also be enhanced by introducing multiple modes of communication including cues such as pointing gestures, speech, and ostensive facial expressions (e.g., raised eye brows). This is because joint attention bids in real-life contexts often involve multiple communicative cues. However, an inevitable challenge associated with increasing the complexity of virtual interactions is that they will also become increasingly difficult to control experimentally.

The Development of RJA and IJA

In **Chapter 3**, I presented evidence in support of the PDPM's claim that RJA and IJA rely on an integrated frontotemporoparietal neural network in adulthood. However, further research is needed to assess the model's claim that RJA and IJA are primarily supported by posterior-parietal and anterior networks that integrate throughout development to support complex joint attention interactions. To empirically support this claim, studies assessing the neural correlates of joint attention in younger participants are needed. Ideally, longitudinal data would be obtained to determine whether there is a critical point during typical development in which this neural integration occurs. It would also be of interest to elucidate how changes in these neural mechanisms correlate with the emergence of joint attention behaviours throughout typical development. The application of connectivity analyses (cf. Cavallo et al., 2015) would also help elucidate how the interaction of brain regions within these joint attention networks differ during RJA and IJA, and across development.

Joint Attention in Special Populations

In **Chapter 3**, I demonstrated that the Catch-the-Burglar paradigm could be used to measure the neural correlates of RJA and IJA in typical development. In **Chapter 4**, I demonstrated the application of this paradigm to investigate the cognitive mechanisms of RJA and IJA impairments in participants with HFA. It would be useful to apply this paradigm in an fMRI study to investigate the relationship between the neural and cognitive mechanisms of joint attention in individuals across the autism spectrum. The data obtained in such a study could be used to evaluate brain-based explanations for impairments in social information processing in autism, such as the Social Brain Theory (Pelphrey, Shultz, Hudac, & Vander Wyk, 2011).

There may also be scope to use the Catch-the-Burglar paradigm to investigate whether joint attention declines with old age, or is impaired in patients with schizophrenia.

In these special populations, people often experience social isolation which decreases their quality of life (Addington & Addington, 2000; Charles & Carstensen, 2010). The Catch-the-Burglar paradigm could be used to investigate whether impairments in social communication skills are related to experiences of social isolation by providing a direct and objective measure of social cognition. It would also be informative to compare any differences in the nature of joint attention impairments across these populations, and investigate whether these differences in performance are reflected by differences in measures of brain function.

One inevitable challenge of investigating the cognitive and neural mechanisms of joint attention in special populations is accounting for heterogeneity in social information processing abilities. This is particularly pertinent to autism research given that autism is a characteristically heterogeneous disorder (Brock, 2011). This heterogeneity was observed in the study presented in **Chapter 4**, which found that one third of the autism sample were exceptionally slow to establish eye contact on IJA trials while two thirds of the sample were not. Future work is needed to better discriminate between individuals with autism who are more likely than others to experience certain types of joint attention deficits. By identifying individuals with difficulties in a particular aspect of social interaction, we will be able to better determine how to best cater their needs, whether through our everyday interactions with them, or through the development of intervention programs focused on training social information processing in interactive contexts.

Autism Intervention

Quantitative and qualitative data from the post-experimental interviews conducted in the study outlined in **Chapter 4** suggested that second person virtual reality paradigms, such as the Catch-the-Burglar paradigm, may provide an effective means for delivering social communication intervention programs to individuals with autism. The individuals with autism interviewed in my study indicated a stronger preference for engaging in a

virtual social interaction over a real face-to-face interaction. Some commented that the virtual interface provided a "filter" in which they could become accustomed to one social cue at a time (e.g., gaze cues). Others explained that the virtual interface felt less intimidating or demanding than real social interactions, but at the same time felt genuinely engaging and social. Second person paradigms, such as the Catch-the-Burglar paradigm, may allow individuals with autism to engage in a hedonic social interaction without the anxiety and stress that may be associated with daily interactions (cf. Kandalaft, Didehbani, Krawczyk, Allen, & Chapman, 2013). The flexibility of virtual interfaces also allows the simulated interaction to be made increasingly complex so that individuals can generalise new social communication skills in their daily lives. Future studies might explore whether such a paradigm can provide an effective training context for individuals with autism.

Separating Cognitive and Hedonic Components of EAJA

In **Chapter 5**, I presented evidence that the centro-parietal P350 ERP was larger to incongruent responses made by a partner following a self-initiated bid for joint attention than congruent responses. This enhanced P350 could have resulted from the cognitive processes of evaluating the attentional perspective of the virtual partner, or it may reflect the affective response associated with achieving or avoiding joint attention. The latter interpretation stems from fMRI studies that have found hedonic-related increased activation in the ventral striatum when a social partner's gaze shift resulted in the achievement of joint attention rather than the avoidance of joint attention (Pfeiffer et al., 2014; Schilbach et al., 2010). In order to examine the specific influence of affective processes on the P350 in joint attention contexts, future studies could investigate whether the P350 response changes when the achievement of joint attention is inconsistent with the participant's goals. For example, this could be achieved by measuring the P350 response when joint attention experiences are evaluated during competitive interactions where the participant must locate a visual target (e.g., a burglar) without revealing the target's

location to their opponent. If the P350 effect is the same under competitive and collaborative contexts, it would suggest that this ERP reflects the evaluation of joint attention achievement irrespective of its affective outcome. This manipulation could also be applied to fMRI studies to elucidate whether the activity in regions associated with reward processing (i.e., ventral striatum; McClure, York, & Montague, 2004) is intrinsic to the achievement of joint attention or if it depends on the context of the interaction (Schilbach et al., 2010).

Conclusions

The studies in this dissertation have made several important methodological and theoretical contributions to the field of social neuroscience and autism research. First, I have extended upon existing second person approaches to develop a joint attention paradigm that indexes behavioural and neural responses associated with RJA, IJA and EAJA. Second, I have discovered that RJA and IJA rely on common neural substrates within a right-lateralised frontotemporoparietal network comprising MFG, IFG, MTG, precentral gyrus, pSTS, TPJ and precuneus. Third, I have conducted the first study to provide an objective measure of RJA and IJA behaviour in adults with autism spectrum disorders. Importantly, these behaviours were observed in the context of complex social interactions that closely mimicked real world experiences. I discovered that joint attention difficulties in autism may stem from impaired monitoring of the attention and intentions of a social partner rather than impaired gaze orienting reflexes. Fourth, I found in adults with typical development that the brain discriminates the outcome of a self-initiated joint attention bid approximately 350 ms after observing a responsive gaze shift. Finally, I found that in order to effectively simulate an ecologically valid interaction, participants must believe that the social stimulus they interact with (e.g., virtual character) represents the live actions and intentions of another human being. I also demonstrated that this belief can be established with minimal deception. I hope that these findings will guide future studies

attempting to investigate the cognitive and neural mechanisms of social communication in typical development, the causes of their impairment in autism, and how interventions can be designed to support social information processing and communication in individuals on the autism spectrum.

References

Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. A. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of Autism & Developmental Disorders, 39*(1), 84-96. doi: 10.1007/s10803-008-0601-7

Addington, J., & Addington, D. (2000). Neurocognitive and social functioning in schizophrenia: a 2.5 year follow-up study. *Schizophrenia Research, 44*(1), 47-56. doi: http://dx.doi.org/10.1016/S0920-9964(99)00160-7

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

Brock, J. (2011). Commentary: Complementary approaches to the developmental cognitive neuroscience of autism – reflections on Pelphrey et al. (2011). *Journal of Child Psychology and Psychiatry, 52*(6), 645-646. doi: 10.1111/j.1469-7610.2011.02414.x

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

Caruana, N., de Lissa, P., & McArthur, G. (2015). The neural time course of evaluating self-initiated joint attention bids. *Brain and Cognition*. doi: http://dx.doi.org/10.1016/j.bandc.2015.06.001

Cavallo, A., Lungu, O., Becchio, C., Ansuini, C., Rustichini, A., & Fadiga, L. (2015). When gaze opens the channel for communication: Integrative role of IFG and MPFC. *NeuroImage, 119*, 63-69. doi: http://dx.doi.org/10.1016/j.neuroimage.2015.06.025

Charles, S., & Carstensen, L. L. (2010). Social and emotional aging. *Annual Review of Psychology, 61*, 383-409. doi: 10.1146/annurev.psych.093008.100448

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions Royal Society London Biological Sciences, 358*, 315-324. doi: 10.1098/rstb.2002.1199

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. *Developmental Psychology, 40*(2), 271-283. doi: http://dx.doi.org/10.1037/0012-1649.40.2.271

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. doi: http://dx.doi.org/10.1016/j.neuroimage.2009.01.063

Kandalaft, M., Didehbani, N., Krawczyk, D., Allen, T., & Chapman, S. (2013). Virtual reality social cognition training for young adults with high-functioning autism. *Journal of Autism and Developmental Disorders, 43*(1), 34-44. doi: 10.1007/s10803-012-1544-6

Kincade, 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. *Journal of Neuroscience, 25*(18), 4593-4604.

McClure, S. M., York, M. K., & Montague, P. R. (2004). The neural substrates of reward processing in humans: The modern role of fMRI. *The Neuroscientist, 10*(3), 260-268. doi: 10.1177/1073858404263526

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20*(1), 115-128. doi: 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 Research, 2*(1), 2-21. doi: 10.1002/aur.61

Murray, D. S., Creaghead, N. A., Manning-Courtney, P., Shear, P. K., Bean, J., & Prendeville, J.-A. (2008). The relationship between joint attention and language in children with autism spectrum disorders. *Focus on Autism & Other Developmental Disabilities, 23*(1), 5-14. doi: 10.1177/1088357607311443

Patterson, T. L., Moscona, S., McKibbin, C. L., Davidson, K., & Jeste, D. V. (2001). Social skills performance assessment among older patients with schizophrenia. *Schizophrenia Research, 48*(2–3), 351-360. doi: http://dx.doi.org/10.1016/S0920-9964(00)00109-2

Pelphrey, K. A., Shultz, S., Hudac, C. M., & Vander Wyk, B. C. (2011). Research Review: Constraining heterogeneity: the social brain and its development in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 52*(6), 631-644. doi: 10.1111/j.1469-7610.2010.02349.x

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., & Vogeley, K. (2014). Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *NeuroImage, 101*, 124-137. doi: 10.1016/j.neuroimage.2014.06.061

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience Biobehavioural Reviews, 37*(10 Pt 2), 2516-2528. doi:10.1016/j.neubiorev.2013.07.017

Redcay, E., Dodell-Feder, D., Mavros, P. L., Kleiner, A. M., Pearrow, M. J., Triantafyllou, C., Gabrieli, J. D., & Saxe, R. (2012). Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. *Human Brain Mapping, 34,* 2511-2523. doi: 10.1002/hbm.22086.

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. doi: 10.1016/j.neuroimage.2010.01.052

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience, 7*(5), 499-500. doi: http://www.nature.com/neuro/journal/v7/n5/suppinfo/nn1223_S1.html

Saxe, R., Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience, 1*(3-4), 284-298. doi: 10.1080/17470910601000446

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah N. J., Fink, G. R., & Vogeley, K. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience, 22*(12), 2702-2715. doi: 10.1162/jocn.2009.21401.

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. *Nature Reviews Neuroscience, 12*(3), 154-167. doi: 10.1038/nrn2994

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. *Journal of Cognitive Neuroscience, 16*(5), 817-827. doi: 10.1162/089892904970799

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. doi: 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. doi: 10.1371/journal.pone.0094339

Supplementary Materials

Chapter 3

Supplementary Material 1

When you look at one of your houses, the door will open.

Inside you will find the burglar………. or an empty house.

Sometimes all of your houses will empty.

This means the burglar is hiding in one of your partner's houses.
Your partner will show you where he is.
You must look in the correct location to catch the burglar.

Sometimes you **will** find the burglar.

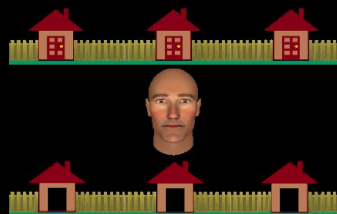You must then show your partner where the burglar is.
Then they can help you catch the burglar.

You and your partner will also complete the same task on your own.
The virtual interface will be turned off. So the avatar's eyes will close.

If you don't find the burglar.
1. Look at the dot
2. Then an arrow will guide you

If you DO find the burglar.
1. Look at the dot
2. Wait for the dot to turn green
3. Look back at the burglar's location

Supplementary Material 2

Table 1.

Neural Correlates for Responding to Joint Attention minus Initiating Joint Attention

(RJA - RJAc) - (IJA-IJAc)

	H	**BA**	**x**	**y**	**z**	**k**	**T**
Precentral Gyrus	L	6	**-32**	**-24**	**66**	40	5.32
* Precuneus	L	3	**-18**	**-32**	**52**	16	3.97

Note. Regions were assigned using SPM5 Anatomy Toolbox. Results are based on an uncorrected threshold of (p<0.005), with an extent threshold of 10 voxels. Coordinates are given in Montreal Neurologic Institute Space. H= Hemisphere, BA=Brodmann Area. *Cluster labels based on nearest grey matter to which the significant cluster extended.

Chapter 4

Supplementary Material 1
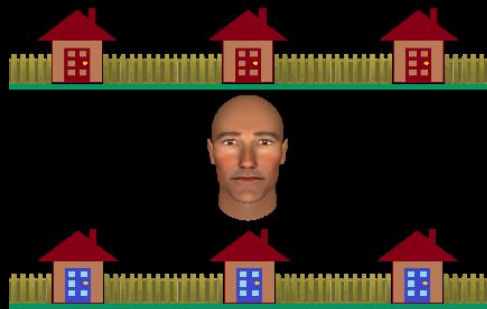
When you look up, the avatar looks up.

Player 2
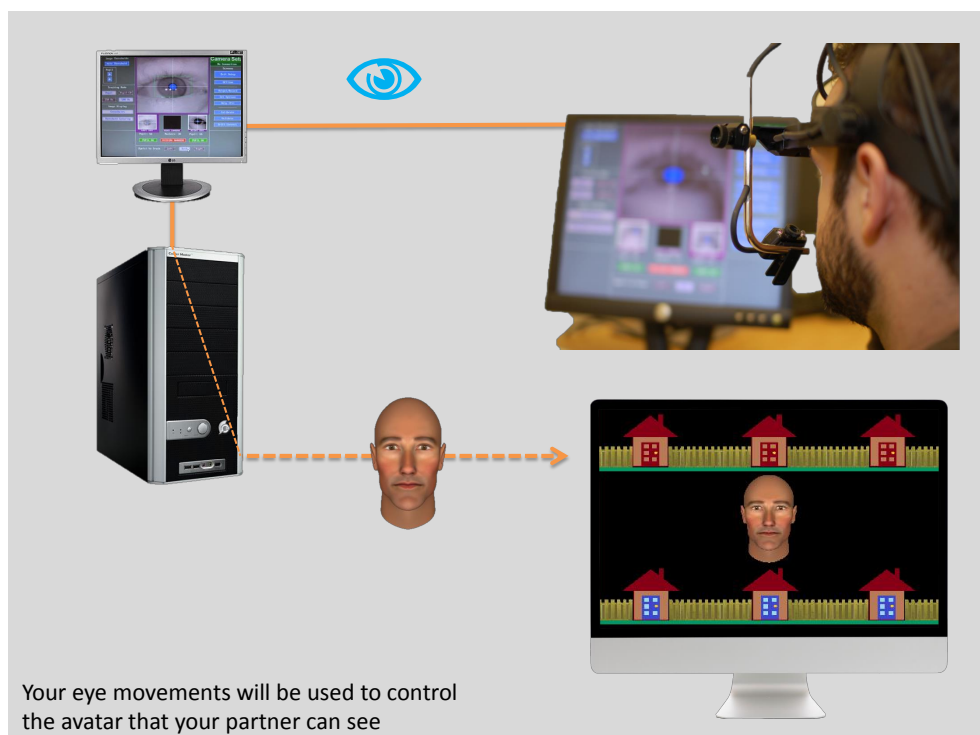
In the same way, your partner will control
the avatar that you see.
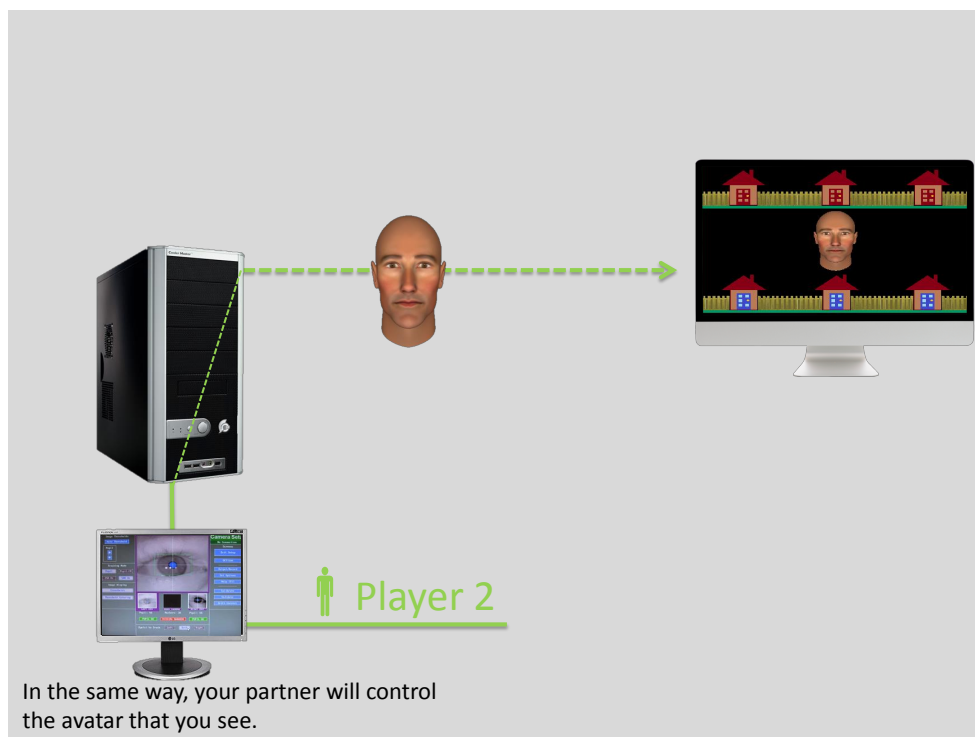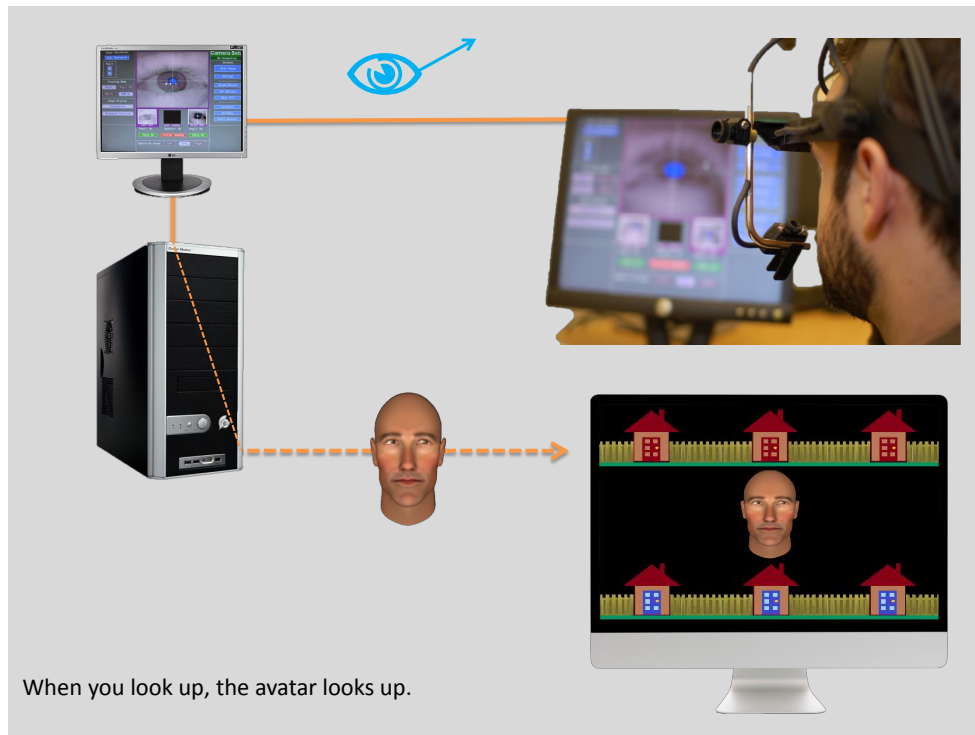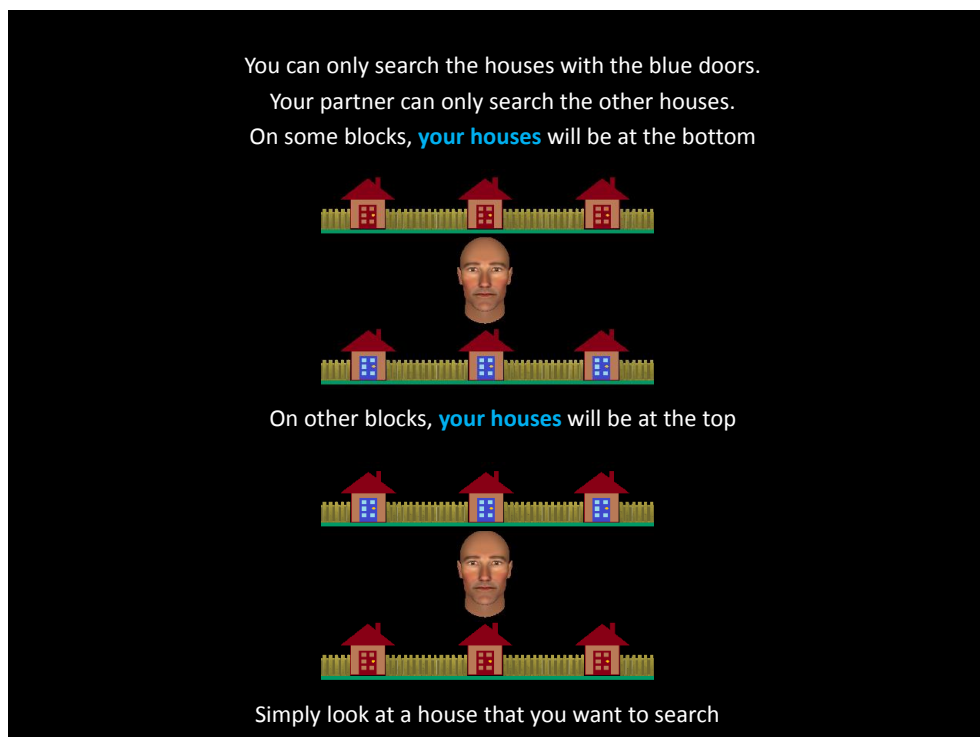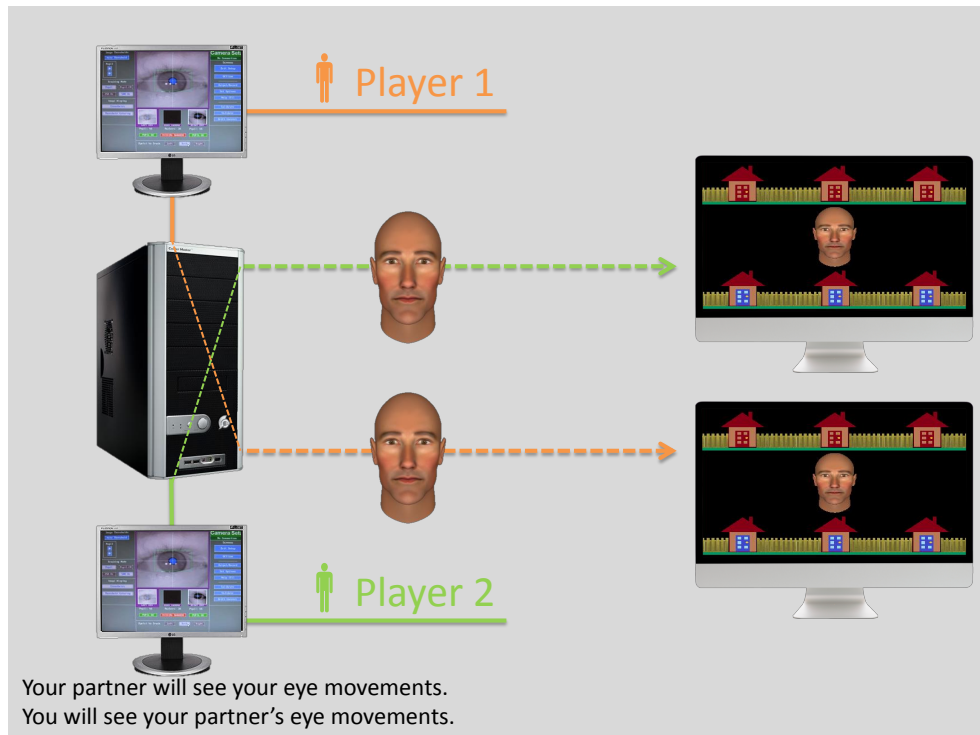
Your partner will see your eye movements.
You will see your partner's eye movements.

You can only search the houses with the blue doors.
Your partner can only search the other houses.
On some blocks, **your houses** will be at the bottom

On other blocks, **your houses** will be at the top

Simply look at a house that you want to search

When you look at one of your houses, the door will open.
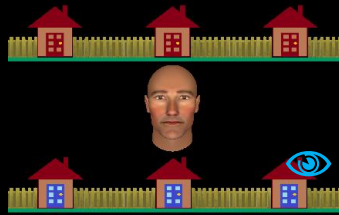
Inside you will find the burglar………. or an empty house
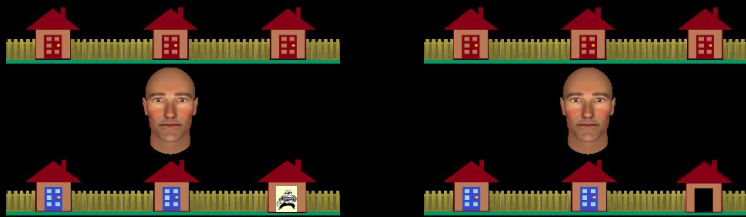
Sometimes all your houses will empty
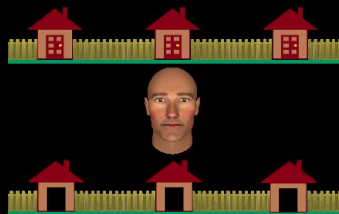
This means the burglar is hiding in one of your partner's houses.
Your partner will show you where the burglar is.
You must look in the correct location to catch the burglar.
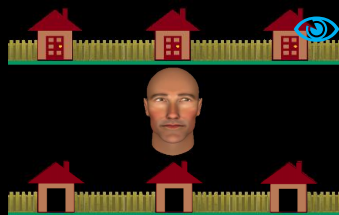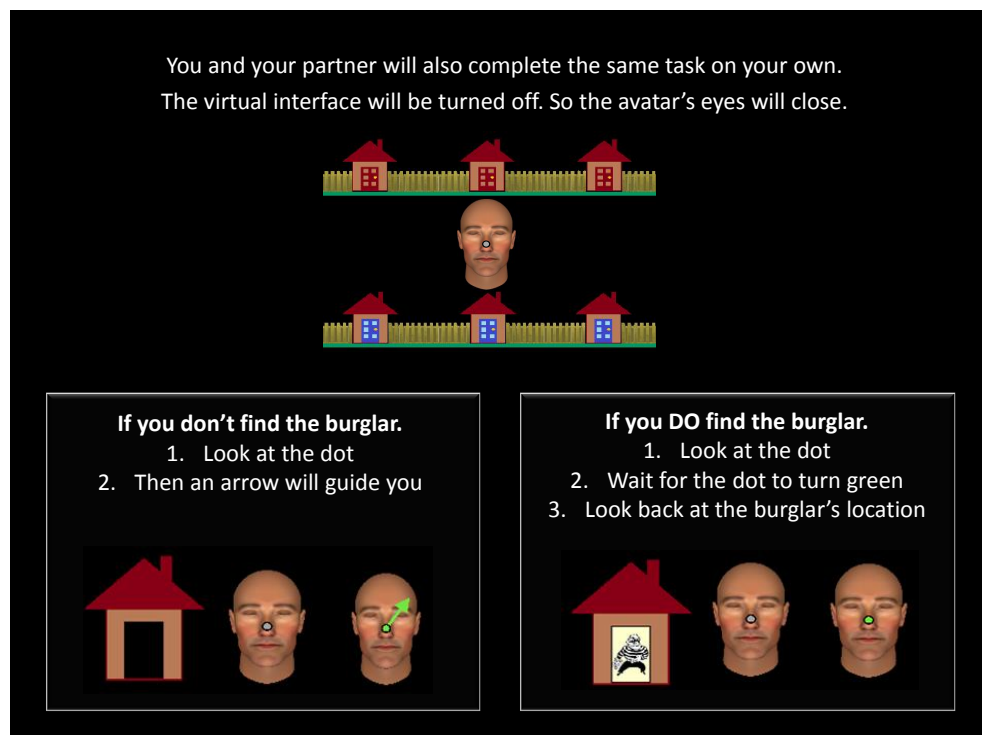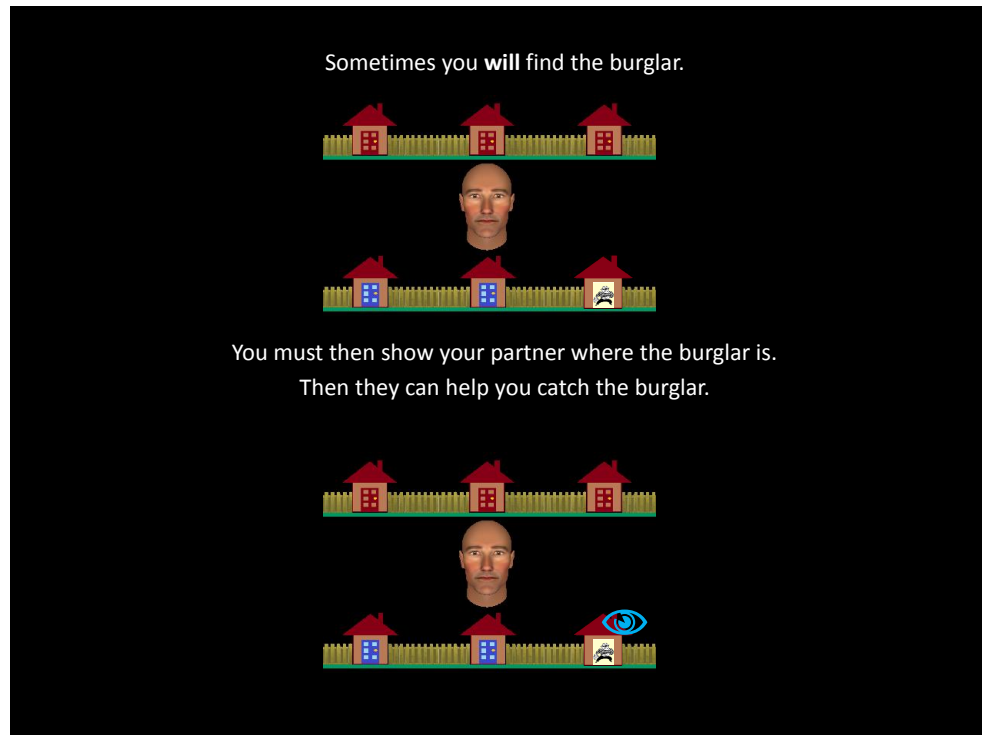
Sometimes you **will** find the burglar.

You must then show your partner where the burglar is.
Then they can help you catch the burglar.

You and your partner will also complete the same task on your own.
The virtual interface will be turned off. So the avatar's eyes will close.

If you don't find the burglar.
1. Look at the dot
2. Then an arrow will guide you

If you DO find the burglar.
1. Look at the dot
2. Wait for the dot to turn green
3. Look back at the burglar's location

Supplementary Material 2

Table 1.

Full ANOVA output.

Source	df	SS	MS	F	p	η_p^2
RJA Accuracy						
Condition	32.00	1422.06	1422.06	6.64	.015	0.17
Block	32.00	338.07	338.07	3.15	.086	0.09
Group	1.00	4005.08	4005.08	9.06	.005	0.22
Condition * Group	1.00	1117.64	1117.64	5.22	.029	0.14
Block * Group	1.00	0.00	0.00	0.00	.999	0.00
Condition * Block	32.00	615.36	615.36	9.53	.004	0.23
Condition * Block * Group	32.00	61.21	61.21	0.95	.338	0.03
RJA Saccadic RT						
Condition	32.00	2005859.40	2005859.40	86.74	.000	0.73
Block	32.00	231122.75	231122.75	11.23	.002	0.26
Group	1.00	259961.24	259961.24	3.67	.064	0.10
Condition * Group	1.00	90329.83	90329.83	3.91	.057	0.11
Block * Group	1.00	98890.84	98890.84	4.80	.036	0.13
Condition * Block	32.00	77062.31	77062.31	5.10	.031	0.14
Condition * Block * Group	32.00	67849.98	67849.98	4.49	.042	0.12
IJA Accuracy						
Condition	32.00	41.94	41.94	1.50	.230	0.04
Block	32.00	40.88	40.88	1.30	.263	0.04
Group	1.00	281.87	281.87	7.04	.012	0.18
Condition * Group	1.00	60.80	60.80	2.17	.150	0.06
Block * Group	1.00	0.00	0.00	0.00	.999	0.00
Condition * Block	32.00	14.84	14.84	0.84	.366	0.03
Condition * Block * Group	32.00	26.76	26.76	1.51	.227	0.05
Dwell Time						
Condition	32.00	3713094.16	3713094.16	7.73	.009	0.19
Block	32.00	2315353.71	2315353.71	19.14	.000	0.37
Group	1.00	2681220.41	2681220.41	2.73	.108	0.08
Condition * Group	1.00	1324200.17	1324200.17	2.76	.107	0.08
Block * Group	1.00	452739.20	452739.20	3.74	.062	0.10
Condition * Block	32.00	633369.73	633369.73	7.55	.010	0.19
Condition * Block * Group	32.00	171834.90	171834.90	2.05	.162	0.06
Premature IJA						
Condition	32.00	452.24	452.24	19.84	.000	0.38
Block	32.00	148.26	148.26	12.99	.001	0.29
Group	1.00	40.26	40.26	0.86	.360	0.03
Condition * Group	1.00	9.53	9.53	0.42	.522	0.01
Block * Group	1.00	3.56	3.56	0.31	.580	0.01
Condition * Block	32.00	2.94	2.94	0.34	.561	0.01
Condition * Block * Group	32.00	4.24	4.24	0.50	.486	0.02

Published Manuscripts

Chapter 3

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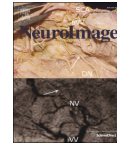
A frontotemporoparietal network common to initiating and responding to joint attention bids

Nathan Caruana *, Jon Brock, Alexandra Woolgar

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

A R T I C L E I N F O

A B S T R A C T

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

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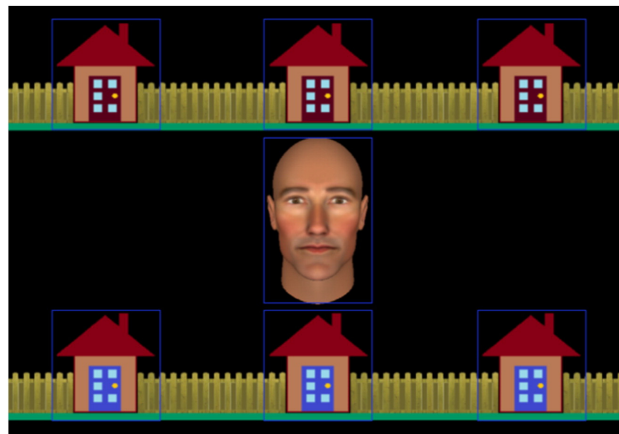
Fig. 1. Gaze areas of interest (GAOIs) overlaid on participants' view of stimuli, represented as blue rectangles.

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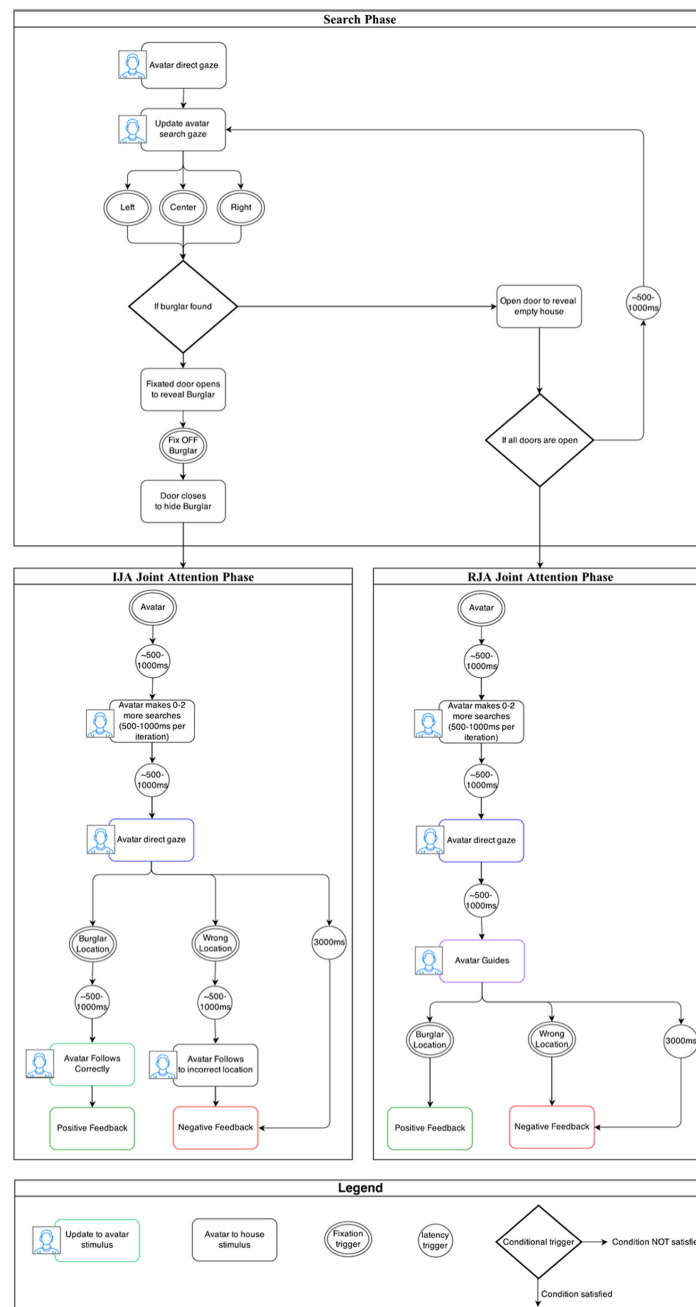
Fig. 2. Schematic summary of interactive algorithm. This algorithm was the same for the social and non-social conditions, apart from the central stimulus that is used (e.g., avatar direct gaze vs green fixation point). This diagram has been labelled to reflect the stimuli in the social conditions.

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 = 64^2; voxel size = $3.0 \times 3.0 \times 3.75$ 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 = $1.0 \times 1.0 \times 1.5$ 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 RJA 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 GAOIs 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 themself, 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_JA 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 sé, 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
Frontal																												
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
	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	−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	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														
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
	R	45	46	24	8	455	6.73																					
Superior frontal gyrus[a]								L	6	−22	6	60	2387	8.98								B	6	20	6	58	12	4.03
																						R	6	18	12	68	58	8.22
																						B	6	4	8	58	42	4.39
SMA																						L	6	−36	−4	50	301	5.65
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														
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	−30	−2	58	2387	8.07														
Temporal																												
Inferior temporal gyrus								R	37	40	−52	−14	27	4.50								R	37	56	−58	−4	1471	6.61
Middle temporal gyrus[a]	R	20	52	−22	−14	52	5.04	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																					
Temporal pole																						L	38	−30	6	−26	14	4.43
Fusiform gyrus								R	V4	38	−60	−12	27	4.45														
Parietal																												
Posterior cingulate cortex								B	29	8	−40	14	36	5.61														
Precuneus	B	7	4	−62	58	355	5.18	B		10	−62	51	5137	8.15	B	5	2	−52	54	93	4.06	B		10	−54	46	80	5.48
Occipital																												
Cuneus								B	18	4	−84	24	5137	8.57														
Calcarine gyrus								B	17	0	−96	−2	13	4.41								B	18	−4	−80	14	6118	11.66
Lingual gyrus								L	19	−22	−66	−4	64	6.58														
Superior occipital gyrus								R		18	−82	32	5137	7.36														
Middle occipital gyrus								L	V5	−46	−74	0	5137	7.60								R		30	−74	30	6118	9.59
																						L	V5	−46	−74	0	6118	8.83
Subcortical																												
Thalamus								R		12	−12	6	187	9.88								R		12	−12	6	21	5.13
Amygdala[a]	B	34	16	−6	−10	21	4.67																					
Cerebellum								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.

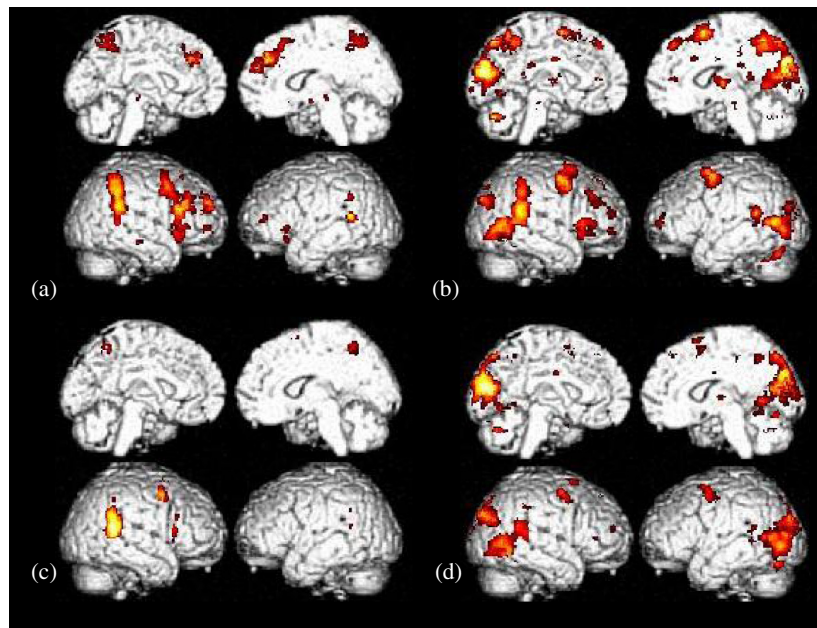*N. Caruana et al. / NeuroImage 108 (2015) 34–46*

Fig. 4. Thresholded statistical parametric maps showing activity associated with (a) Responding to joint attention (RJA − RJAc) (b) Initiating joint attention (IJA − IJAc) (c) Activation common to Responding and Initiating (d) Initiating over and above Responding [(IJA − IJAc) − (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 − IJAc)].

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 − IJAc), IJA resulted in significantly more activation in a bilateral frontotemporoparietal network than RJA; (IJA −IJAc) − (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 > IJAc 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 (IJAc) 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 with 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.

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., Csmaioni, 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.
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.UK 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.
Kincade, 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

46   *N. Caruana et al. / NeuroImage 108 (2015) 34–46*

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/suppinfo/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.

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 perfromance. 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.
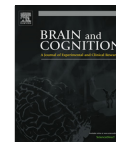
Chapter 5

ELSEVIER

CrossMark

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

A B S T R A C T

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.

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 × 12° 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.
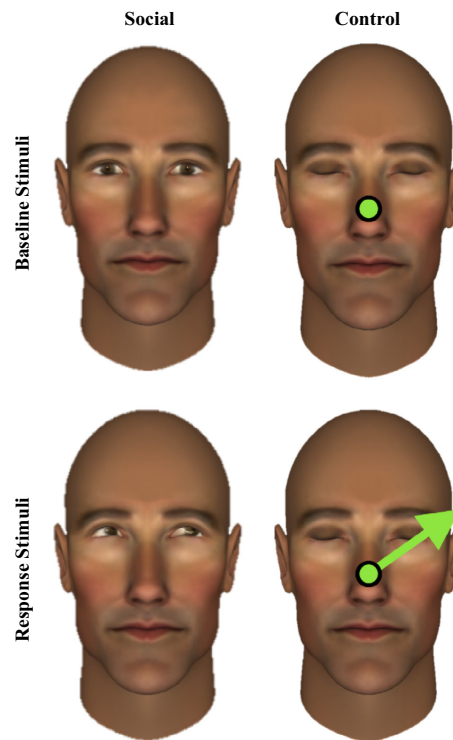

Fig. 1. Central stimuli used in the social and control versions of the gaze-contingent task.

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 counterbalanced across all acquisition blocks, and thus the virtual character was equally likely to gaze towards one of the three houses not
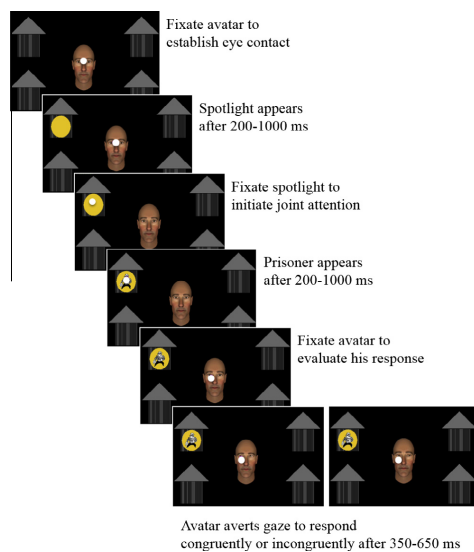
Fixate avatar to establish eye contact

Spotlight appears after 200-1000 ms

Fixate spotlight to initiate joint attention

Prisoner appears after 200-1000 ms

Fixate avatar to evaluate his response

Avatar averts gaze to respond congruently or incongruently after 350-650 ms

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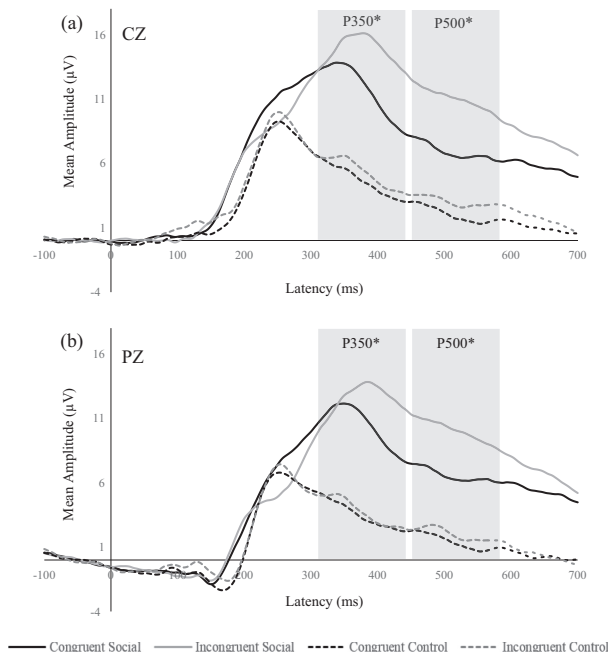
Fig. 3. Group average waveforms comprising the P350 and P500 at (a) Cz and (b) Pz electrodes. Epochs were time-locked to the onset of the virtual character's gaze shift. * Indicates a significant Group * Condition interaction effect.

$SD = 20.95$; 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

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

	CZ		PZ	
	Congruent	Incongruent	Congruent	Incongruent
P350 mean amplitude				
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)
P500 mean amplitude				
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
N170 peak amplitude				
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*).

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'

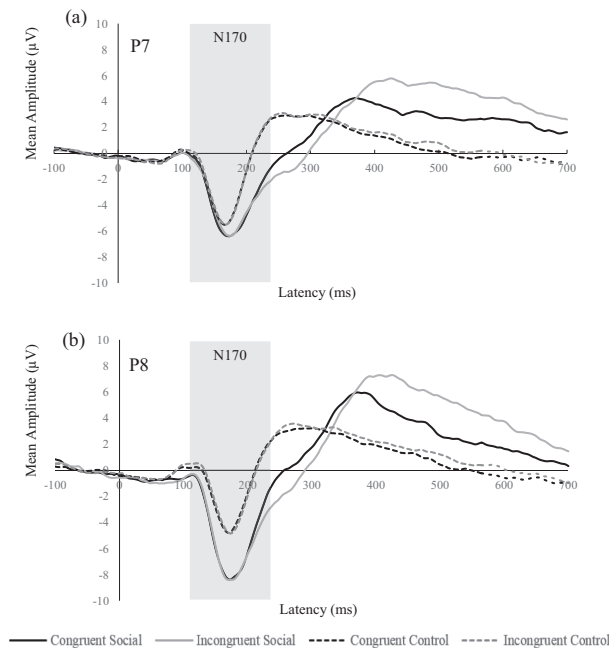*N. Caruana et al. / Brain and Cognition 98 (2015) 43–52*

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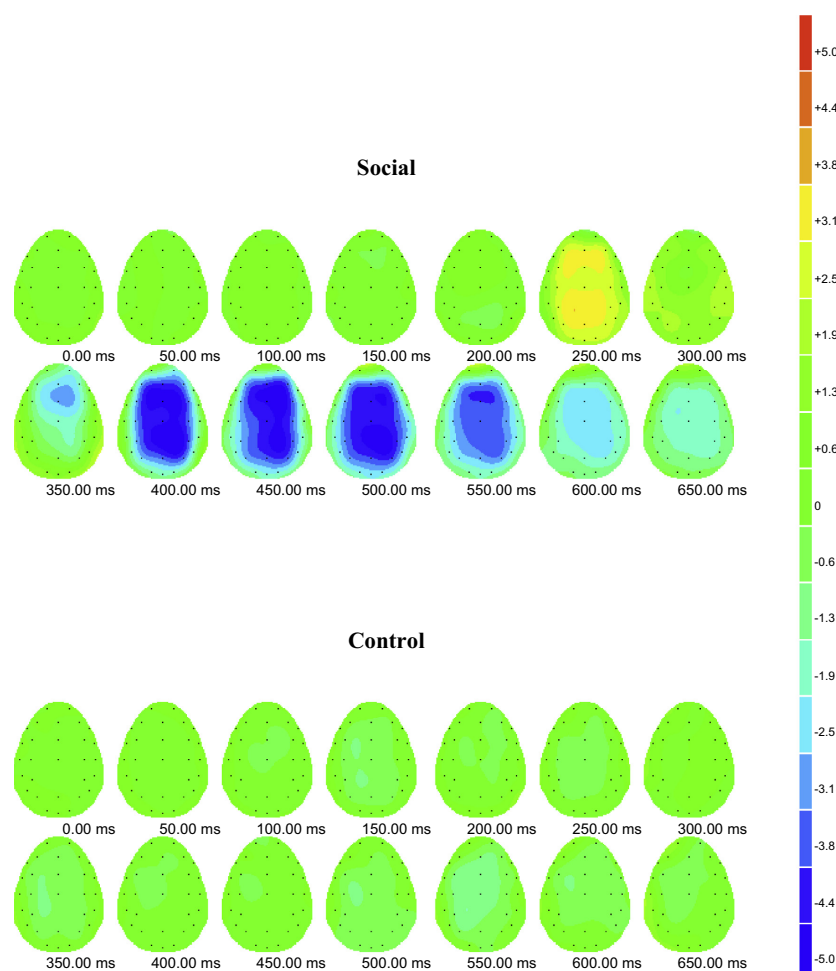
Fig. 5. Effect topographies (congruent–incongruent) by group condition.

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.

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.

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)*. Berkley.

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.

Ethics Approvals

Ethics application ref: 5201200021 - Amendment Approved 🟡 PROJECT/Ethics - PhD

Ethics Secretariat ethics.secretariat@mq.edu.au via hotmail.com
to Mr, Dr ▾

Dear Nathan

Re: Brain mechanisms of attention and social cognition in autism (Ref: 5201200021)

Thank you for your email and SONA advertisement. The following amendment has been approved:

1. The addition of Associate Professor Genevieve McArthur and Dr Peter De Lissa as Associate Investigators on the project.

2. The addition of FMRI (functional magnetic resonance imaging) and EEG (electroencephalography) scanning techniques to the project in addition to the approved MEG scanning technique. The techniques involving MEG, EEG and fMRI will be used in different experiments. Participants will not be asked to complete the tasks using all three of the imaging techniques, unless they apply to participate in these different studies separately. At this stage non-clinical participants will be tested for the fMRI and EEG studies.

3. The duration of the testing session for fMRI experiments has increased to 1 hour. The testing session using EEG will be 45 minutes.

4. In the fMRI study structural abnormalities might be detected during the acquisition process. If an MRI suggests signs of a potential abnormality, a neuroradiologist will be asked to review the data and advice on whether the participant should be referred on to see their doctor for diagnostic testing. The risks are detailed in the information and consent form.

5. The information and consent forms have been amended to reflect the above changes.

6. Neurotypical adult participants will be recruited from the SONA system and will be paid for their participation.

Please do not hesitate to contact the Ethics Secretariat if you have any questions or concerns.

Kind regards
Fran

Ms Fran Thorp
Human Research Ethics Officer

Office of the Deputy Vice Chancellor (Research)

Ethics Secretariat
Research Office
Level 3, Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia
T: +61 2 9850 6848

MACQUARIE UNIVERSITY

Office of the Deputy Vice-Chancellor (Research)
Research Office
C5C Research HUB East, Level 3, Room 324
MACQUARIE UNIVERSITY NSW 2109 AUSTRALIA

Phone   +61 (0)2 9850 7850
Fax   +61 (0)2 9850 4465
Email   ethics.secretariat@mq.edu.au

22 September 2014

Dr Jon Brock
Department of Cognitive Science
Faculty of Human Sciences
Macquarie University   2109

Dear Dr Brock,

Reference No: 5201200021

Title: Brain mechanisms of attention and social cognition in autism

Thank you for your correspondence dated 19 September 2014 submitting an amendment request to the above study. Your proposed amendment was reviewed and approved by the HREC (Medical Sciences) Ethics Secretariat.

I am pleased to advise that ethical approval of the following amendments to the above study has been granted:

1. Collect data in collaboration with researchers at the University of Western Australia and Curtin University, Perth.

2. Addition of the following people to the project personnel:
- Dr Nadine Kloth (UWA)
- A/Prof Romina Palermo (UWA)
- Dr Heidi Stieglitz Ham (Curtin)
- Dr Sonya Girdler (Curtin)

Approval of this amendment is on the condition of ethical approval from the above universities.

Details of this approval are as follows:

The following documentation submitted with your email correspondence has been reviewed and approved by the HREC (Medical Sciences):

Documents reviewed	Version no.	Date
Macquarie University HREC Request for Amendment Form	2.0	Received 19/09/2014
Participant Information & Consent Form entitled *A Study of Attention processing and Perspective Taking in Autism*		

1

Human Ethics
Office of Research Enterprise
The University of Western Australia
M459, 35 Stirling Highway
Crawley WA 6009 Australia
T   +61 8 6488 4703 / 3703
F   +61 8 6488 8775
E   humanethics@uwa.edu.au
CRICOS Provider Code: 00126G

Our Ref: RA/4/1/7185

22 October 2014

Ms Romina Palermo
School of Psychology
MBDP: M304

Dear Ms Palermo

HUMAN RESEARCH ETHICS OFFICE – RECOGNITION OF ETHICS APPROVAL FROM ANOTHER HUMAN RESEARCH ETHICS COMMITTEE

Project: Brain Mechanisms of Attention & Social Cognition in Autism - Recogntion Macquarie University HREC Approval 5201200021

Thank you for your correspondence enclosing the necessary documents to facilitate recognition of the ethics approval for the above project granted by an external Human Research Ethics Committee (HREC) registered with the National Health and Medical Research Council (NHMRC).

It is noted that you have ethics approval from Macquarie University Ethics Review Committee (Human Research, approval number 5201200021.

The UWA students and researchers identified as working on this project are:

UWA Researchers:

Name	Faculty / School	Role
Ms Romina Palermo	School of Psychology	Chief Investigator
Dr Nadine Kloth	School of Psychology	Co-Investigator
Dr Jonathan Brock	Macquarie University	Co-Investigator

Student(s): *None specified*

Although The University of Western Australia reserves the right to subject any research involving its staff and students to its own ethics review process, in this case, the Human Research Ethics Office has recognised the existing approval of the external HREC. The project is exempt from ethics review at UWA and the involvement of the above-listed researchers has been authorised. Any conditions for the recognition of the external HREC's existing approval are listed below:

Special Conditions

    1. *Approving HREC to receive annual reports, amendments and notification of adverse events*

You are reminded that it will be the responsibility of the approving HREC to ensure compliance with all ethics requirements and to monitor and report on the project. However, should any relevant ethics issues arise during the course of the project, you should inform the Human Research Ethics Office of The University of Western Australia.

If you have any queries, please contact the HEO at humanethics@uwa.edu.au.

Please ensure that you quote the file reference – RA/4/1/7185 – and the associated project title in all future correspondence.

Yours sincerely