



## Meta-analysis

# Simulating social interactions for the experimental investigation of joint attention



Nathan Caruana<sup>a,b,c,\*</sup>, Genevieve McArthur<sup>a,b,d</sup>, Alexandra Woolgar<sup>a,b,c</sup>, Jon Brock<sup>a,b,d,e</sup>

<sup>a</sup> Department of Cognitive Science, Macquarie University, Sydney, Australia

<sup>b</sup> ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, Australia

<sup>c</sup> Perception in Action Research Centre, Macquarie University, Sydney, Australia

<sup>d</sup> Centre for Atypical Neurodevelopment, Macquarie University, Sydney, Australia

<sup>e</sup> Department of Psychology, Macquarie University, Sydney, Australia

## ARTICLE INFO

## Article history:

Received 15 August 2016

Received in revised form

13 December 2016

Accepted 20 December 2016

Available online 24 December 2016

## Keywords:

Joint attention

Eye gaze

Eye-tracking

Social interaction

Neuroimaging

## ABSTRACT

Social interactions are, by their nature, dynamic and reciprocal – your behaviour affects my behaviour, which affects your behaviour in return. However, until recently, the field of social cognitive neuroscience has been dominated by paradigms in which participants passively observe social stimuli from a detached “third person” perspective. Here we consider the unique conceptual and methodological challenges involved in adopting a “second person” approach whereby social cognitive mechanisms and their neural correlates are investigated within social interactions (Schilbach et al., 2013). The key question for researchers is how to distil a complex, intentional interaction between two individuals into a tightly controlled and replicable experimental paradigm. We explore these issues within the context of recent investigations of joint attention – the ability to coordinate a common focus of attention with another person. We review pioneering neurophysiology and eye-tracking studies that have begun to address these issues; offer recommendations for the optimal design and implementation of interactive tasks, and discuss the broader implications of interactive approaches for social cognitive neuroscience.

© 2016 Elsevier Ltd. All rights reserved.

## Contents

1. fMRI studies of joint attention .....	116
1.1. Responding to joint attention bids .....	116
1.2. Initiating joint attention bids .....	120
1.3. Evaluating the achievement of joint attention (EJA) .....	120
2. Critical issues affecting a second-person measurement of joint attention .....	121
2.1. Complex interactions .....	121
2.2. Intentional interactions .....	121
2.3. Engaging interactions .....	122
2.4. Controlled measurement of social and non-Social cognition .....	123
3. Conclusions .....	123
Acknowledgements .....	124
References .....	124

\* Corresponding author at: Department of Cognitive Science, 16 University Ave, Macquarie University, Sydney, NSW 2109, Australia.

E-mail address: [nathan.caruana@mq.edu.au](mailto:nathan.caruana@mq.edu.au) (N. Caruana).

Humans are innately social creatures with a biological imperative for social interaction (Baumeister and Leary, 1995). We seek social interactions to share information, to accomplish shared goals, and to enjoy shared interests. As social cognitive neuroscientists, our aim is to understand the cognitive and neural mechanisms that underlie these vital social behaviours, their emergence during development, and the ways in which they may diverge from the norm in conditions such as autism, schizophrenia, and various forms of acquired or degenerative brain injury. Until recently, research in this field has relied on paradigms in which participants are presented with social stimuli (e.g., faces or videos of social interactions) that they view and respond to from a detached “third person” perspective. However, as Schilbach et al. (2013) have cogently argued, the cognitive and neural mechanisms involved in completing such tasks are not necessarily the same as those engaged in everyday social interactions where individuals must process information from a “second person” (i.e., you and I) perspective embedded *within* the interaction. Accordingly, the challenge for social cognitive neuroscientists is to develop interactive paradigms that achieve this ecological validity, whilst at the same time maintaining close experimental control. At the forefront of such efforts have been recent studies (including our own) investigating the neural correlates of joint attention. Our objective in this paper is to extract the key lessons from this nascent field of research and draw out the broader implications for social cognitive neuroscience.

The term “joint attention” refers to our ability to simultaneously coordinate attention between a social partner and an object or event of interest (Bruner, 1974). In a typical joint attention episode, 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 third component as evaluating the achievement of joint attention (EAJA).<sup>1</sup> These behaviours emerge during reciprocal and ongoing social interactions, and are greater than (or at least different to) the combined behaviours of each individual acting alone (Hobson, 2008). As such, joint attention can *only* be experienced from within a face-to-face interaction involving at least two people. Therefore, a “second person” approach is essential to the measurement and investigation of joint attention.

To date, most research on joint attention has been conducted by developmental psychologists. It has been established that infants begin to display RJA behaviours at approximately six months of age when they reflexively follow the gaze of others around them (Mundy et al., 1994). Initiating behaviours appear somewhat later, typically between six and twelve months of age (Mundy et al., 1994). These emerging abilities are considered to be a key component of children’s social and cognitive development, playing a crucial role in language development, and learning in general (Adamson et al., 2009; Baron-Cohen, 1995; Charman, 2003; Mundy et al., 1990, 2009; Murray et al., 2008; Tomasello, 1995). For instance, if a parent describes or names an object whilst directing an infant’s attention to that object, and the infant responds by attending to the same object, then he or she has an opportunity to form associations between the visual, lexical, and semantic repre-

sentations of the object (Baldwin, 2014). Furthermore, elay in the development of joint attention is strongly associated with autism spectrum disorders. It is one of the earliest recognisable symptoms of the condition (Lord et al., 2000) and reliably predicts the severity of social and linguistic impairments that autistic children experience (Charman, 2003; Dawson et al., 2004; Lord et al., 2000; Mundy et al., 1990; Stone et al., 1997).

Yet despite its importance in both typical and atypical development, little is currently known about the underlying cognitive and neural mechanisms that support joint attention. Models of joint attention have been proposed (e.g., Baron-Cohen, 1995; Mundy et al., 2009), but these are largely descriptive, lack detail, and are yet to be rigorously tested. The superficial nature of our current understanding is due, at least in part, to the inherent challenges in creating adequate experimental measures of joint attention. Standardized observational protocols, such as the Early Social Communication Scales (Mundy et al., 2003) and the Autism Diagnostic Observation Schedule (Lord et al., 2000), can reliably measure joint attention behaviours in young children; however, these scales do not allow for the experimental manipulations or large number of trials necessary for investigating the underlying cognitive or neural mechanisms. Until recently, experimental studies of “joint attention” were largely restricted to variations of the Posner-cueing paradigm (Posner, 1980) in which response times to a visual target are influenced by the image of a pair of eyes looking either towards or in the opposite direction of the target. However, this paradigm taps low-level “reflexive” orienting of attention (Friesen and Kingstone, 1998) and studies of autistic individuals have failed to find consistent evidence of impairment, in contrast to the findings from more naturalistic measures (Leekam, 2015; Nation and Penny, 2008). The challenge, therefore, is to develop controlled experimental tests that capture the intentional, mutual, and communicative aspects of joint attention.

Taking up this challenge, in 2010, researchers in the United States, Japan and Germany independently published three studies that effectively kick-started the field of second person neuroscience (Redcay et al., 2010; Saito et al., 2010; Schilbach et al., 2010). Each study used functional magnetic resonance imaging (fMRI) to investigate the neural correlates of joint attention. Subsequent fMRI studies (including our own) have built on and refined the methodological innovations in these three pioneering studies. In the following section of this paper, we review this growing body of research. Our focus is less on the particular findings of these studies (see Pfeiffer et al., 2013 for a comprehensive review) and more upon the tasks themselves. In particular, we consider how the three components of joint attention (RJA, IJA, and EAJA) have been operationalised and (critical to any fMRI study) the baseline conditions with which they have been contrasted.

In the second half of the paper, we provide a synthesis of the critical issues affecting the measurement of joint attention using a second-person approach. In particular, we consider the importance of realistically complex interactions, the intentional nature of the interaction, and the question of whether participants need to interact with (or believe they are interacting with) a real person. In addition to the fMRI studies, we include insights from recent eye-tracking and electroencephalography (EEG) studies that have addressed these questions directly. We conclude by considering directions for future research.

## 1. fMRI studies of joint attention

### 1.1. Responding to joint attention bids

As noted above, the field of second person neuroscience arguably began with three fMRI studies of joint attention published

<sup>1</sup> 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, in the existing experimental literature most researchers have continued to use the term “joint attention” when describing EAJA during social interactions (cf. Pfeiffer et al., 2013).

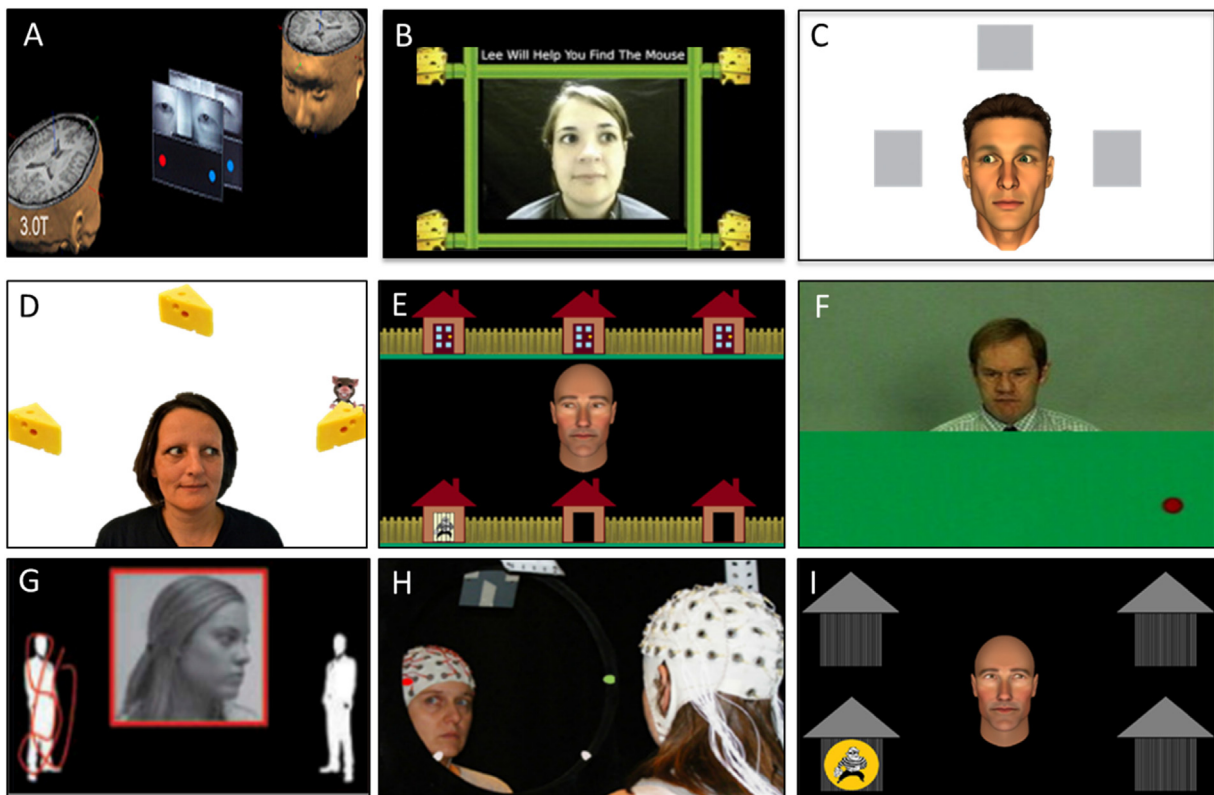
**Table 1**  
Summary of experimental tasks using interactive methods to investigate the neural mechanisms of joint attention.

Authors	Participants	Stimulus	Component	Task	Baseline	Contrast
Caruana et al., 2015a (see also Caruana et al. 2017)	TD adults	Face of gaze-contingent avatar (believed to be controlled by another participant).	RJA	Participant and avatar search houses on screen for burglar. The participant finds all his/her houses empty, waits for avatar to make eye-contact, then follows avatar's gaze to catch the burglar.	Avatar's eyes remain closed. The participant finds all his/her houses empty. A central arrow cues the burglar's location.	Baseline controls for eye-movement and attention demands but cues differ perceptually.
			IJA	Participant and avatar search houses on screen for burglar. The participant finds burglar in one of his/her houses, waits for avatar to make eye-contact, then looks back to burglar location. The avatar follows.	Avatar's eyes remain closed. The participant finds burglar in one of his/her houses, looks back at avatar, waits for fixation point to turn green, looks back to burglar location.	Baseline controls for eye-movement and attention demands but cues differ perceptually.
Caruana et al., 2015b, 2016a	TD adults	Gaze-contingent avatar	EAJA	Participants initiated joint attention towards one of four "prison exists" to inform their partner of a prison break in a collaborative game. The partner responded by looking at the same location to achieve joint attention. Participants initiated joint attention by fixating one of two human silhouettes. Their "partner" turned their head to look at the same silhouette.	Participants initiated joint attention towards one of four "prison exists" to inform their partner of a prison break in a collaborative game. The partner responded by looking at a different location to avoid joint attention. Participants initiated joint attention by fixating one of two human silhouettes. Their "partner" turned their head to look at the other silhouette.	Stimuli and participant's behaviour matched across conditions.
Gordon et al., 2013	TD adults	Gaze-contingent video	EAJA	Participants initiated joint attention by fixating one of two human silhouettes. Their "partner" turned their head to look at the same silhouette.	Participants initiated joint attention by fixating one of two human silhouettes. Their "partner" turned their head to look at the other silhouette.	Stimuli and participant's behaviour matched across conditions.
Lachat et al., 2012	TD adults	Pairs of participants sit facing each other through circular window with LEDs around frame	EAJA (intentional)	One participant looks towards LED and the other follows them.	One participant looks towards LED and other participant looks to different/opposite location.	Baseline introduces antisaccade element.
			EAJA (incidental)	Both participants look towards the same LED.	Both participants look towards a different LED.	Stimuli and participant's behaviour matched across conditions.
Oberwelling et al., 2016	TD children and adolescents	Photo of face (own mother or stranger) that changes in response to participant's gaze.	RJA	Participant follows gaze of photo to look at one of three cheeses.	Participant ignores photo's gaze and attends to peripheral colour cue.	Conditions matched for eye-movements but differ in attention to central vs peripheral cues.

Table 1 (Continued)

Authors	Participants	Stimulus	Component	Task	Baseline	Contrast
			EAJA	Participant fixates on one of three cheeses. Photo “responds” by looking at the same cheese.	Participant fixates on one of three cheeses. Photo “responds” by shifting its gaze downwards and away from the cheeses.	Contrast reveals EAJA rather than IJA (as authors claim).
Pfeiffer et al., 2014	TD adults	Face of gaze-contingent avatar (believed to be controlled by another participant on some trials, and by a computer on others).	EAJA	Fixate on one of two squares. Avatar programmed to look at the same square.	Fixate on one of two squares. Avatar programmed to look at opposite square.	Contrast reveals EAJA. Authors also examined the effect of manipulating the proportion of congruent avatar responses within each block of trials.
Redcay et al., 2010, 2012	TD and ASD adults	Live video feed showing partner’s full face.	RJA	Follow partner’s gaze to catch a mouse hiding behind one of four cheeses.	Search for mouse hiding behind one of four cheeses while partner’s eyes remain closed.	Baseline removes social element but requires attention to peripheral rather than central cues.
			IJA	Search for mouse hiding behind one of four cheeses. Partner follows to same location.	Search for mouse hiding behind one of four cheeses while partner’s eyes remain closed.	Baseline controls for eye-movements. Unclear if participants process partner’s behaviour during the task.
Saito et al., 2010	TD adults	Live video feed showing partner’s eyes.	RJA	Follow partner’s gaze to look at one of two circles presented on the screen.	Avoid joint attention by looking at the circle in the opposite direction to their partner’s gaze.	Baseline controls for eye-movements but subtracts away gaze processing and introduces antisaccade component.
Schilbach et al., 2010	TD adults	Face of gaze-contingent avatar (believed to be controlled by another participant).	RJA	Follow avatar’s gaze to look at one of three squares.	Look at a different square to the avatar. <sup>a</sup>	Baseline controls for eye-movements but subtracts away gaze processing and introduces antisaccade component.
			EAJA	Fixate on one of three squares. Avatar programmed to look at the same square.	Fixate on one of three squares. Avatar programmed to look at different square. <sup>a</sup>	Contrast reveals EAJA rather than IJA (as authors claim).
Williams et al., 2005	TD adults	Pre-recorded video	EAJA	Participants looked at a red ball presented on one side of the screen and their “partner” looked at the same location.	Participants looked at a red ball presented on one side of the screen and their “partner” looked in the opposite direction.	Stimuli and participant’s behaviour matched across conditions.

<sup>a</sup> Schilbach et al. also included a further baseline condition in which participants passively viewed the same visual display but with the avatar’s eyes closed. This was subtracted from each of the other four conditions prior to examining the contrasts of interest described in the table.



**Fig. 1.** Still frames from interactive joint attention tasks: (a) Saito et al., 2010; (b) Redcay et al., 2010; (c) Schilbach et al., 2010; (d) Oberwelland et al., 2016; (e) Caruana et al., 2015a, 2017; (f) Williams et al., 2005; (g) Gordon et al., 2013; (h) Lachat et al., 2012; (i) Caruana et al., 2015b. All images reused with permission.

in 2010 (Redcay et al., 2010, 2012; Saito et al., 2010; Schilbach et al., 2010). In each case, participants in the scanner interacted with a partner using eye gaze cues. Saito et al. (2010) employed a “hyperscanning” design in which two participants were scanned simultaneously in separate MRI systems as they interacted via live video feed (see Fig. 1, Panel A; Table 1). Each participant could see their partner’s eyes at the top of the screen. At the bottom of the screen 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 (RJA) and look at the circle in the same location.

Redcay et al. (2010); 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 viewed via a live video feed. The task was presented as a collaborative game in which the participant and Lee had to catch a mouse that was concealed behind one of four cheeses located in each corner of the screen (see Fig. 1, Panel B; Table 1). 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.

In Schilbach et al.’s (2010) study, instead of a live video feed, participants interacted with an anthropomorphic virtual character (i.e., an avatar; see Fig. 1, Panel C; Table 1) whom they believed was controlled by a confederate outside the scanner via an eye-tracking device. In reality, the virtual character was controlled by an algorithm that was contingent on the participant’s own eye movements, supporting reciprocity in the interaction whilst maintaining control over the virtual partner’s behaviour (Wilms et al., 2010). On RJA trials (referred to as OTHER.JA), the virtual character looked towards one of three squares and participants were required to respond congruently by following his gaze. Recently, Oberwelland et al. (2016) employed a similar task in a study of joint attention in children and

adolescents. The avatar was replaced by a photographic face (either a stranger or the participant’s own mother) (Fig. 1, Panel D; Table 1) and participants were made aware that they were interacting with a computer rather than a real person.

Our own recent fMRI study (Caruana et al., 2015a) built on these seminal investigations. Following Schilbach et al. (2010), participants interacted via an eye-tracker with a virtual character whom they believed to be controlled by a real person (“Alan”), but was actually controlled by a gaze-contingent algorithm. Following Redcay et al. (2012), the participant and “Alan” engaged in a cooperative game in which they were required to both fixate upon a target location in order to catch a burglar hiding in one of six houses (see Fig. 1, Panel E; Table 1). The novel feature of our task was that, rather than being instructed explicitly to initiate or respond to joint attention, the participants’ social role became apparent implicitly during the trial. Specifically, each trial began with a search phase in which participants searched their allotted houses by fixating on each house in turn to reveal its contents (to them only). On RJA trials, participants found all of their allotted houses to be empty. They were then required to wait for Alan to complete his own search, make eye contact, and then guide them to the burglar’s location.

Each of these studies identified RJA-related activity in brain regions associated with social cognition in their samples of typically-developed individuals. These include the medial prefrontal cortex (Caruana et al., 2015a; Redcay et al., 2012; Schilbach et al., 2010), posterior superior temporal sulcus (Caruana et al., 2015a; Redcay et al., 2012), temporoparietal junction (Caruana et al., 2015a; Redcay et al., 2012), intraparietal sulcus (Saito et al., 2010), occipital gyrus (Oberwelland et al., 2016), precuneus, insula, and amygdala (Caruana et al., 2015a). Of note, however, is that despite essentially investigating the same phenomenon, there is relatively little agreement or overlap in the precise brain regions identified across studies.

These discrepancies may in part reflect different choices of baseline condition. Schilbach et al. (2010) and Saito et al. (2010) contrasted their RJA trials with trials in which participants were instructed to actively avoid joint attention by looking in the opposite or different direction to their partner. As the baseline also involved processing the gaze of the social partner, subtracting this from RJA trials may have removed activation related to gaze processing. It may have also introduced additional attention orienting, action inhibition and oculomotor demands associated with making an “antisaccade” away from the cued location (Pfeiffer et al., 2013). The other three studies have contrasted RJA with responses to a non-social spatial cue. Oberwilling et al. (2016) required participants to ignore the gaze cue and instead look towards one of the cheeses, which changed colour. In Redcay et al.’s study, participants could see the mouse’s tail and were required to “catch” it themselves while their partner’s eyes remained closed. The virtual character’s eyes also remained closed in our own study (Caruana et al., 2015a), but a green arrow indicated the burglar’s location in lieu of a gaze cue. In each case, participants were required to make the same pattern of eye-movements in the baseline condition as in the RJA condition, thereby controlling for many of the attention orienting and oculomotor control demands of the RJA task. However, there are inevitable differences in the visual stimuli presented to participants.

### 1.2. Initiating joint attention bids

Four of the fMRI studies described above have investigated IJA in addition to Schilbach et al. (2010) instructed participants to fixate on one of three squares located around their virtual partner’s face. The partner would then respond congruently to achieve joint attention (referred to as SELF.JA) or, in the baseline condition (referred to as SELF.NOJA) avoid joint attention by saccading to a different square. Similarly, Oberwilling et al. (2016) instructed participants to fixate on one of three cheeses located around a human face. The face would either look at the same cheese (referred to as JA-Self) or look downward and away from the three cheeses (referred to as Control-Self). Therefore, in both studies, participants performed exactly the same task in IJA and baseline conditions and any differences in brain activity reflected the different outcomes of their IJA bids. Thus, although both studies involved an IJA condition, it is more appropriate to discuss them in the following section that focuses on EAJA.

In the IJA condition of Redcay et al.’s (2012) study, participants were instructed to locate a mouse whose tail was protruding from behind one of four cheeses. Joint attention was initiated by fixating upon the relevant cheese, whereupon their partner, Lee, interacting via video link, would follow to the same location. In the baseline condition participants located the mouse on their own while Lee’s eyes remained closed, thus controlling for activation associated with performing the visual search and making eye movements. A potential concern with Redcay et al.’s task is that participants knew their role as an “initiator” (or “responder”) before each trial began. In IJA blocks, participants could complete the IJA task whilst ignoring Lee, knowing that she would follow them regardless of whether or not they communicated their intention to initiate a joint attention bid.

In our Catch-the-Burglar task, IJA trials were signalled when the participant located the burglar in one of their allotted houses. They then had to saccade back to the virtual character, wait for him to complete his own search and, having established eye contact, guide him to the burglar location. The avatar did not respond unless eye contact had been established, thereby ensuring that the participant had to interact with him in order to complete the task. Again, we compared IJA to a closely matched baseline condition designed to control for the non-social components of IJA (e.g., attention ori-

enting, action inhibition and oculomotor control). Following the search phase, participants fixated upon a small grey circle placed between the virtual character’s closed eyes. When this turned green (analogous to the avatar making eye contact), the participant was required to saccade back to the correct location (as they would do if they were guiding Alan).

The results of our study and those of Redcay et al. (2010); Redcay et al. (2012) are in fact remarkably consistent. Both studies associated IJA with brain regions previously implicated in social cognition including the middle frontal gyrus, inferior frontal gyrus, precentral gyrus, middle temporal gyrus, posterior superior temporal gyrus, temporoparietal junction, and precuneus. Of note, both studies also directly contrasted brain responses in IJA and RJA conditions, reporting increased IJA-related activation of the right inferior frontal gyrus – a region widely implicated in inhibitory control of planned actions (e.g., Aron et al., 2004). Inspection of our eye-tracking data suggested a possible explanation: Participants made a surprisingly large number of “premature” saccades, failing to wait for eye contact with their partner before making saccade to initiate joint attention at the burglar location. This suggests that an important and heretofore ignored component of IJA is the requirement to withhold the initiating “action” until the participant is sure they have the attention of their partner.

### 1.3. Evaluating the achievement of joint attention (EAJA)

Evaluating the achievement of joint attention (EAJA) is another important component of joint attention as it helps guide future joint attention behaviours if initial attempts to achieve joint attention with another person have been unsuccessful. A number of studies have investigated the neural correlates of EAJA using gaze congruency tasks, where participants are required to fixate upon an object and their social partner looks at the same object, thereby achieving joint attention, or responds incongruently by looking at another location to avoid joint attention.

An early study by Williams et al. (2005) used a non-interactive task to investigate the incidental achievement of joint attention. Participants were required to follow a red dot that moved between the bottom-left and bottom-right of the screen. The top half of the display showed a video of a man either looking at the dot (joint attention) or toward a different location (no joint attention; see Fig. 1, Panel F; Table 1). Contrasting these two conditions revealed activation in the ventral medial prefrontal cortex, a region associated with representing the mental perspectives of others (Amodio and Frith, 2006).

More recently, Gordon et al. (2013) used an interactive eye-tracking task to investigate EAJA. Participants were presented with a display that comprised a central video frame depicting the upper torso and face of a female named “Sally”. This was flanked by two human silhouettes (see Fig. 1, Panel G; Table 1). On each trial, participants fixated on one of the silhouettes in order to “guide” Sally. The video of Sally was then played to depict her turning her head to look at either the same or opposite location. Congruent trials were associated with increased activation of the striatum, anterior cingulate, right fusiform gyrus, right amygdala, and parahippocampus.

As noted earlier, Schilbach et al. (2010) and Oberwilling et al.’s (2016) fMRI studies also effectively investigated EAJA within the context of their IJA tasks. Recall that participants looked towards one of three targets and the virtual partners either responded to achieve joint attention by looking at the same target or avoided joint attention by looking elsewhere (see Fig. 1, Panels C and D; Table 1). In Schilbach et al.’s study, participants showed increased activation of the ventral striatum for congruent trials. Using the same task, Pfeiffer et al. (2014) grouped trials into short blocks that varied in the proportion of joint attention trials and, consistent with Schilbach et al.’s results, found that blocks with a high proportion of

joint attention trials were associated with greater activation in ventral striatum. However, Oberwelland et al. failed to find evidence of striatal activation in their study of younger children and adolescents. Instead, gaze congruency was associated with activation in anterior prefrontal cortex and left anterior insular.

As with RJA and IJA, there is no consistent pattern of results across studies. Nonetheless, a number of studies have reported striatal activity associated with EAJA (Gordon et al., 2013; Pfeiffer et al., 2014; Schilbach et al., 2010), which may reflect the hedonic value of successfully achieving joint attention (cf. Craig, 2009; McClure et al., 2004). Williams et al. (2005) found no such effect, although joint attention was achieved incidentally rather than deliberately. These studies have also differed in whether participants believed they were interacting with another living human agent. Further studies are required to determine whether these methodological differences can account for some of their discrepant findings.

## 2. Critical issues affecting a second-person measurement of joint attention

A second person approach to social cognition and neuroscience demands both ecological validity and experimental control. However, as the preceding review demonstrates, there is considerable tension between these two requirements. In the remainder of this paper we consider developments and directions 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, are genuinely engaging, and maintain experimental control. These methodological considerations are relevant to the investigation of joint attention as well as the application of second person approaches to social cognition and neuroscience research in general.

### 2.1. Complex interactions

A core aim of the second person approach is to measure social cognition while people participate in dynamic social interactions that simulate the complexity of everyday experiences (Schilbach et al., 2013). However, experimental measures that have attempted to distil elements of joint attention behaviour within simulated interactions do not necessarily capture this complexity. The Posner-cueing tasks using gaze stimuli described in the Introduction are a clear example, but even the more interactive joint attention tasks tend to isolate the initiating or responding “event” from the dynamic social context that they would naturally occur in.

Of particular concern is that social cues in such tasks are often unambiguous, and are displayed by social partners whose behaviour is highly predictable. For example, in RJA experiments, participants are typically required to respond to a single gaze cue that they know to be communicative (e.g., Oberwelland et al., 2016; Redcay et al., 2012; Saito et al., 2010; Schilbach et al., 2010). This is unlike genuine social interactions where a social partner is constantly making eye movements, and only a subset of these signal communicative joint attention bids. Similarly, in IJA experiments, participants have been asked to make a single eye-movement towards a particular location with full knowledge that their partner will follow (e.g., Redcay et al., 2012). Again, this is unlike real social interactions, where we must ensure that we have our partner’s attention, and convey our intention to communicate with them before attempting to initiate a joint attention bid. Thus, whilst these tasks provide close experimental control and allow a large number of relatively brief trials, they do not capture a number of cognitive processes that would typically be used to navigate com-

plex social interactions, including decisions about (1) whether one’s social partner intends to communicate, (2) which gaze shifts are communicative, and (3) whether to respond to, or initiate, a joint attention bid.

As outlined above, the search phase of our interactive Catch-the-Burglar task addresses some of these issues. Rather than overtly instructing participants to engage in RJA or IJA on each trial (cf. Oberwelland et al., 2016; Redcay et al., 2012; Saito et al., 2010; Schilbach et al., 2010), participants intuitively determined their role depending on whether they found the burglar (IJA trials) or not (RJA trials). On RJA trials, the participant then had to differentiate between gaze shifts displayed by their virtual partner that were intentionally guiding them to the burglar, and those that were being made as he completed his search. Likewise, on IJA trials, the participant could not assume that their partner would follow every gaze shift they made. Rather, they had to establish eye contact before engaging in IJA. Using eye contact in this way is a critical aspect of *intentional* gaze-based joint attention experiences as it informs the initiator that they have their partner’s attention, and signals the initiator’s intent to communicate to the responder (Cary, 1978; Senju and Johnson, 2009; Tylén et al., 2012).

The importance of differentiating between communicative and non-communicative gaze shifts is underscored by a recent study that we conducted outside the scanner (Caruana, McArthur, Woolgar and Brock, 2017). We compared eye-movement response times in the RJA condition of our Catch-the-Burglar task with response times in a modified version of the task that removed the search component at the beginning of each trial. We predicted that response times would be shorter in this “NoSearch” condition because, despite the fact that participants must respond to the exact same gaze cue, they would not have to make any decision as to whether or not they should follow the cue. In contrast, removing the search component from the control condition should have no effect because participants respond to an unambiguous arrow cue in both conditions. Both predictions were supported, indicating that the original Catch-the-Burglar task did indeed capture the RJA processes of identifying a social partner’s intention to communicate, which pre-empts social communication (cf. Cary, 1978) and thus adaptive RJA behaviour.

Our task captures at least some of the complexity of everyday social interactions, but more can be done. For instance, all the experimental studies reviewed here 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 and head movements) and these joint attention cues all occur against a background of multiple competing social and non-social drivers of attention. Future studies could investigate how social cues from multiple modalities are simultaneously used during joint attention episodes. The challenge for researchers will be to maintain experimental control as the social context becomes increasingly complex. Recent developments in immersive 3-D virtual reality may offer unique opportunities in this regard (Georgescu et al., 2014).

### 2.2. Intentional interactions

Another core aim of the second person approach is to develop paradigms in which social cognition is engaged in the same way as it would be in an everyday social interaction. Typically, during a joint attention episode, the initiator has a goal in mind, such as requesting assistance, sharing information, or communicating an interest. In such cases, joint attention is achieved when the initiator intentionally guides the other person to attend to a particular object or location. It has been argued, therefore, that joint attention should be studied within a social context where the initiator’s joint

attention bid is intentional, and the responder perceives it as such (Tomasello, 1995).

The fMRI studies reviewed above vary considerably in terms of how joint attention is achieved. Redcay et al. (2012) introduced the notion of embedding the joint attention episode in the context of a cooperative game, asking participants to use eye gaze to communicate information that was relevant to their current goal (i.e., catching a mouse). We have adopted a similar approach with our Catch-the-Burglar task. Both tasks are somewhat unnatural insofar as the goal is achieved simply by the act of the responder looking in the correct location. A more ecologically valid scenario might see participants using eye gaze as a directional cue but then achieving their goal by giving a manual response, perhaps directing a cursor to the relevant location. Nonetheless, these two tasks do motivate participants to engage in intentional RJA and IJA behaviours. Other studies reviewed above have not provided a purpose or motivation for pursuing a joint attention experience. Participants have either been provided with an instruction to guide or follow their partner's eye gaze (e.g., Gordon et al., 2013; Saito et al., 2010; Schilbach et al., 2010) or have used a task in which participants incidentally find themselves looking at the same location as their partner (Williams et al., 2005).

To our knowledge, only one study has directly investigated the effect of establishing an intentional context for joint attention experiences. Lachat et al. (2012) used electroencephalography (EEG) to measure the neural correlates of EAJA. Participants were tested in pairs, looking at each other through a circular window that had coloured lights positioned around its circumference (Fig. 1, Panel H; Table 1). In the “social instruction” condition, joint attention was achieved intentionally with one participant choosing a light to look at and the other participant following (or in the control condition looking in the opposite direction). In the “colour instruction” condition, participants ignored their partner and looked at a light of a particular colour. Joint attention was achieved incidentally when the two participants were instructed to look at the same location. Lachat et al. found that, regardless of how joint attention was achieved, congruent trials were associated with a power decrease in 11–13 Hz oscillations, suggesting that there are at least some neurocognitive processes that reflect the achievement of joint attention independently of the context in which it occurs.

However, it remains possible that other neural correlates of joint attention may be sensitive to the way in which it is achieved. As noted in the section on EAJA, intentional joint attention experiences appear to be associated with striatal activation (Gordon et al., 2013; Pfeiffer et al., 2014; Schilbach et al., 2010), whereas incidentally achieved joint attention does not (Williams et al., 2005). This may reflect differences in the intrinsic social reward value of the two experiences. However, proper support for this claim requires a direct comparison between intentional and incidental joint attention within the same fMRI experiment. Future studies could also investigate joint attention behaviours in other social contexts. Researchers could contrast joint attention in competitive as opposed to cooperative games, in requesting behaviour, or in learning and teaching scenarios.

### 2.3. Engaging interactions

A third tenet of the second person approach is that social cognition should be studied during genuine social interactions. One approach taken by a number of studies is for two participants to interact with one another either face-to-face (Lachat et al., 2012) or via a live video feed (Redcay et al., 2012; Saito et al., 2010). Some of these studies (e.g., Saito et al., 2010) have involved simultaneous scanning of both partners. One mooted advantage of this ‘hyper-scanning’ approach is that it allows investigation of “inter-brain” effects such as correlated activation in common brain regions of

the two interacting people (i.e., “brain-to-brain coupling”). However, it remains unclear what such findings mean other than that some brain regions are implicated in both responding to and initiating joint attention. Moreover, given that these correlations are necessarily mediated by the *behaviour* of the two participants, a more useful approach may be to measure the correlation between one person's social behaviour and their interactive partner's brain response (i.e., “stimulus-to-brain coupling”; see Konvalinka and Roepstorff, 2012 for an in-depth discussion).

A key limitation of live interaction paradigms (hyper-scanning or otherwise) is that they come at the expense of experimental control over various aspects of the interaction, such as the precise timing of behaviour and the potential display of ostensive facial expressions. To address these concerns, other studies have engaged participants in an interaction with a virtual partner whose behaviour is contingent on the participant's behaviour, but is controlled by a computer algorithm. As a number of authors have noted, virtual reality offers an innovative tool for maintaining experimental control whilst allowing participants to experience the “copresence” that characterises genuine social interactions (Biocca et al., 2003; Georgescu et al., 2014; Wilms et al., 2010). However, an important question with regard to virtual reality is whether or not participants need to believe that their virtual partner represents a living human agent.

Data from two of our own recent studies provide evidence that beliefs about human agency do indeed influence the way participants approach the task and engage with their partner. In one study (Spirou et al., in preparation), participants completed the Catch-the-Burglar task described above but half were deceived into believing that they were interacting with a real person (as per our previous studies), while the other half were informed at the outset that the avatar was entirely controlled by a computer program. Participants in the “human” condition were slower to saccade back to the avatar having found the burglar and were more likely to attempt to guide the avatar before making eye contact than participants in the “computer” condition. The implication is that participants expect a degree of flexibility from a human social partner (that they know not to expect from a computer). In particular, participants sometimes act as if they expect their human partner to simply *know* that they are attempting to guide them even when eye contact is not first established. One possibility is that they are providing their virtual partner with other facial cues which, of course, the eye tracker interface is not sensitive to, or able to convey.

The second study measured participants' event-related potentials (ERPs) as they completed a modified version of the Catch-the-Burglar task (Caruana et al., 2015b, 2016a). Participants again interacted with a virtual character, Alan, whose face was depicted in the centre of the screen (Fig. 1 Panel I; Table 1). They were told that Alan was a prison guard responsible for locking down any prison exits that were breached by escaping prisoners. Participants assumed the role of a watch person patrolling four prison exits, depicted as buildings in each corner of the screen. An IJA episode occurred when participants attempted to guide Alan to an exit that had been breached. As in the EAJA studies reviewed above, Alan only responded congruently on half of the trials. When participants believed that Alan was controlled by a real person, we found that the centro-parietal P350 was sensitive to whether he followed them or not (Caruana et al., 2015b). However, this effect was abolished when participants knew Alan was controlled by a computer algorithm (Caruana et al., 2016a).

We found similar effects of participant belief on the earlier N170 response, which has been associated with the perceptual processing of gaze shifts (see Itier and Batty, 2009 for a review). These findings are consistent with earlier studies showing that neural activity during competitive games is significantly modulated by beliefs about whether they are interacting with a human agent or



a computer (Gallagher et al., 2002; McCabe et al., 2001). Together, they suggest that, when individuals adopt an “intentional stance” towards an agent and believe it to represent a real and rational human being, cognitive processes involved in understanding the mental states of others are recruited, and have a top-down effect on the perceptual processing of social information, such as gaze shifts (Wykowska et al., 2014). An interesting avenue for future research would be to investigate the role that individual differences in anthropomorphism have on the influence of agency beliefs during virtual interactions, as some people may have a greater propensity to evaluate non-human stimuli as if they were genuinely human (Waytz et al., 2010).

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 best 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 appear less robotic (Wilms et al., 2010). In our own studies, we have programmed the virtual character to ensure that his behaviour varies across trials in terms of the number and sequential order of gaze shifts made during his search. For the IJA condition, we also programmed him to follow the participant’s gaze even when the participant guided him to the incorrect location (e.g., Caruana et al., 2015a, 2017).

Another challenge is to maintain the illusion of human agency in tasks that require the social partner to behave in a way that is unexpected. For example, in some studies of EAJA, agency beliefs might be affected by the fact that half of the trials involve the virtual partner inexplicably failing to follow the participant’s gaze. Consistent with this view, Pfeiffer et al. (2014) told participants that the avatar was only human-controlled on some blocks and found that participants rated the avatar as more likely to be human-controlled following blocks with a high proportion of congruent responses. In our ERP studies described above (Caruana et al., 2015b, 2016a), we used a cover story to make the virtual character’s non-responsive behaviour seem plausible. Participants were told that their partner was distracted by a secondary task (monitoring conflicts between inmates at other locations within the prison). Whilst participants rated their partner’s performance on the task poorly, they also rated him as being a highly cooperative partner who was pleasant to interact with.

Cover stories are an essential part of convincing participants that the virtual character is being controlled by a real human. Some studies have employed elaborate deceptions, with participants being introduced to confederates with whom they think they will be interacting (cf. Schilbach et al., 2010; Wilms et al., 2010). However, in our studies, we simply told participants that they were interacting with another person, and provided an explanation of how the virtual interface supposedly worked. The subjective ratings provided by participants after the experiment indicated that the deception was successful without the need for a human confederate. Indeed, upon debrief, participants often express disappointment as they were looking forward to meeting “Alan”. Our experience, then, is that it is surprisingly easy to convince participants that they are interacting with a human. Nonetheless, further work is required to determine which features of the task and cover story are necessary for the deception. This will inform the most ethical, effective, and practical induction of these human agency beliefs when using virtual interaction paradigms.

#### 2.4. Controlled measurement of social and non-social cognition

The issues discussed thus far have been specific to the second person approach. However, a critical issue for any endeavour in cognitive neuroscience is that of experimental control. This is essential

in teasing apart the cognitive and neural mechanisms that underpin everyday social behaviour. An important concern raised in our review of fMRI studies has been the choice of baseline condition. In order to isolate the cognitive or neural mechanisms of joint attention, the principle of “pure insertion” asserts that the two conditions being contrasted should only differ with respect to the cognitive process of interest. Previous fMRI studies of RJA have violated this principle by contrasting RJA with a baseline condition in which participants respond to their partner’s eye gaze cue by shifting their gaze towards an incongruent location (e.g., Oberwilling et al., 2016; Schilbach et al., 2010; see also Lachat et al., 2012). As noted earlier, this subtracts away processes relating to the identification of gaze direction whilst introducing the requirement that participants make an effortful antisaccade to a non-cued location. Consequently, it is difficult to interpret the difference in responses across the experimental and “baseline” conditions.

Our own approach has been to compare behavioural and neural responses during joint attention trials with baseline conditions in which participants completed essentially the same task, making the same pattern of eye-movements and attention shifts, but without any social interaction. Differential behavioural or brain responses can be interpreted as reflecting, in the main, the purely social aspects of completing the task. However, even here, we are faced with a number of methodological and conceptual challenges. For example, replacing an eye gaze cue with a non-social (arrow) cue means a confounding difference in the visual stimuli presented to participants. One option might be to use physically identical stimuli but manipulate the participant’s beliefs about the agency of the social stimuli (e.g., Caruana et al., 2016a). Our experience, however, is that it is difficult to convince participants that an avatar is controlled by another human partner if they have previously completed the same task knowing that the avatar is controlled by a computer program. This necessitates a less powerful between-subjects design where some participants are deceived and others are not.

Contrasting joint attention with some form of non-social baseline condition can help to isolate what is special about social attention, and determine whether impairments of joint attention reflect social or non-social cognitive impairments (e.g., Caruana et al., *in press*). Likewise, contrasting different components of joint attention, such as RJA and IJA, can elucidate more fine-grained differences between these related processes. In our recent fMRI study (Caruana et al., 2015a), we went further by performing a conjunction analysis to determine which brain regions were activated by both RJA and IJA (relative to their non-social baselines). This approach has helped to better characterise the neural mechanisms that are common to RJA and IJA, and are thus likely to be at the core of joint attention ability. Future fMRI research could build on these findings, employing multivariate analyses, such as multi-voxel pattern analysis (e.g., Mur et al., 2009) to determine the nature of information that is encoded in each of these core joint attention regions. For instance, some regions may encode information about the direction of gaze, whilst others may encode whether a gaze shift signals communicative intent. This research will be invaluable in future work attempting to specifically characterise the nature of joint attention impairments in conditions such as autism, at the cognitive and neural level. Indeed, a second person approach has considerable potential to advance our understanding of psychiatric disorders that have core impairments in social interactions (see Schilbach, 2016).

### 3. Conclusions

Social cognition is inherently complex. Successful social interactions rely on a combination of “higher-level” cognitive processes

(e.g., intention monitoring), lower level social perception (e.g., gaze processing), as well as a suite of domain general cognitive faculties (Apperly, 2010). The challenge for social cognitive neuroscientists is to tease apart these component mechanisms and their neural correlates in a manner that simultaneously achieves ecological validity and experimental control.

Consistent with the “second person” approach outlined by Schilbach et al. (2013), we have argued that joint attention should be investigated whilst participants are immersed in a social interaction. Recent studies have made important progress in this regard, with virtual reality paradigms arguably providing the optimal balance between validity and control. However, this field of research is still in its infancy and numerous conceptual and methodological challenges remain. Future studies addressing the issues raised in our article should provide new and important insights into the cognitive and neural processes that support our unique ability to engage, understand, and communicate with other human beings.

## Acknowledgements

**Funding:** This work was supported by the Australian Research Council Centre of Excellence for Cognition and its Disorders [CE110001021] [www.ccd.edu.au](http://www.ccd.edu.au). Dr. Woolgar is a recipient of an Australian Research Council Discovery Early Career Award [DE120100898]. Dr. Brock was the recipient of an Australian Research Council Australian Research Fellowship [DP098466]. Dr. Caruana was the recipient of an Australian Postgraduate Award at Macquarie University.

## References

- Adamson, L.B., Bakeman, R., Deckner, D.F., Ronski, M.A., 2009. Joint engagement and the emergence of language in children with autism and Down syndrome. *J. Autism Dev. Disord.* 39 (1), 84–96, <http://dx.doi.org/10.1007/s10803-008-0601-7>.
- Amadio, 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>.
- Apperly, I., 2010. *Mindreaders: The Cognitive Basis of Theory of Mind*. Psychology Press.
- Aron, A.R., Robbins, T.W., Poldrack, R.A., 2004. Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8 (4), 170–177, <http://dx.doi.org/10.1016/j.tics.2004.02.010>.
- Baldwin, D.A., 2014. *Understanding the link between joint attention and language*. In: Moore, C., Dunham, P. (Eds.), *Joint Attention: Its Origins and Role in Development*. Taylor & Francis.
- Baron-Cohen, S., 1995. *Mindblindness*. MIT Press, Cambridge, MA.
- Baumeister, R.F., Leary, M.R., 1995. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* 117 (3), 497–529, <http://dx.doi.org/10.1037/0033-2909.117.3.497>.
- Biocca, F., Harms, C., Burgoon, J.K., 2003. Toward a more robust theory and measure of social presence: review and suggested criteria. *Presence: Teleoperators Virtual Environ.* 12 (5), 456–480, <http://dx.doi.org/10.1162/105474603322761270>.
- Bruner, J.S., 1974. From communication to language: a psychological perspective. *Cognition* 3 (3), 255–287, [http://dx.doi.org/10.1016/0010-0277\(74\)90012-2](http://dx.doi.org/10.1016/0010-0277(74)90012-2).
- Caruana, N., Brock, J., Woolgar, A., 2015a. A frontotemporo-parietal 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>.
- Caruana, N., de Lissa, P., McArthur, G., 2015b. The neural time course of evaluating self-initiated joint attention bids. *Brain Cogn.*, <http://dx.doi.org/10.1016/j.bandc.2015.06.001>.
- Caruana, N., de Lissa, P., McArthur, G., 2016a. Beliefs about human agency influence the neural processing of gaze during joint attention. *Soc. Neurosci.*, <http://dx.doi.org/10.1080/17470919.2016.1160953>.
- Caruana, N., Stieglitz Ham, H., Brock, J., Woolgar, A., Kloth, N., Palermo, R., McArthur, G., 2016b. *Joint attention difficulties in autistic adults: An interactive eye-tracking study*. *Autism* (in press).
- Caruana, N., McArthur, G., Woolgar, A., Brock, J., 2017. Detecting communicative intent in a computerised test of joint attention. *Peer*, <https://doi.org/10.7717/peerj.2899j5:e2899>.
- Cary, M.S., 1978. The role of gaze in the initiation of conversation. *Soc. Psychol.* 41 (3), 269–271, <http://dx.doi.org/10.2307/3033565>.
- Charman, T., 2003. Why is joint attention a pivotal skill in autism? *Philos. Trans. R. Soc. Lond. Biol. Sci.* 358, 315–324, <http://dx.doi.org/10.1098/rstb.2002.1199>.
- Craig, D.B., 2009. *How do you feel now? The anterior insula and human awareness*. *Nat. Rev. Neurosci.* 10 (1), 59–70.
- 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, <http://dx.doi.org/10.1037/0012-1649.40.2.271>.
- Emery, N.J., 2000. The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24 (6), 581–604, [http://dx.doi.org/10.1016/S0149-7634\(00\)00025-7](http://dx.doi.org/10.1016/S0149-7634(00)00025-7).
- Friesen, C., Kingstone, A., 1998. The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon. Bull. Rev.* 5 (3), 490–495, <http://dx.doi.org/10.3758/bf03208827>.
- 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, <http://dx.doi.org/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. *Front. Hum. Neurosci.*, 8, <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. *Soc. Neurosci.*, 1–11, <http://dx.doi.org/10.1080/17470919.2013.832374>.
- Hobson, R.P., 2008. Interpersonally situated cognition. *Int. J. Philos. Stud.* 16 (3), 377–397, <http://dx.doi.org/10.1080/09672550802113300>.
- Itier, R.J., Batty, M., 2009. Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33 (6), 843–863, <http://dx.doi.org/10.1016/j.neubiorev.2009.02.004>.
- Konvalinka, I., Roepstorff, A., 2012. The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci.* 6, 215, <http://dx.doi.org/10.3389/fnhum.2012.00215> <http://journal.frontiersin.org/article/10.3389/fnhum.2012.00215>.
- Lachat, F., Hugueville, L., Lemarechal, J.D., Conty, L., George, N., 2012. Oscillatory brain correlates of live joint attention: a dual-EEG study. *Front. Hum. Neurosci.* 6, 156, <http://dx.doi.org/10.3389/fnhum.2012.00156>.
- Leekam, S., 2015. *Social cognitive impairment and autism: what are we trying to explain?* *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 371 (1686).
- Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., Pickles, A., Rutter, M., 2000. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* 30 (3), 205–223, <http://dx.doi.org/10.1023/A:1005592401947>.
- McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T., 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci.* 98 (20), 11832–11835, <http://dx.doi.org/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. *Neuroscientist* 10 (3), 260–268, <http://dx.doi.org/10.1177/1073858404263526>.
- 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., Sigman, M., Kasari, C., 1994. Joint attention, developmental level, and symptom presentation in autism. *Dev. Psychopathol.* 6 (03), 389–401, <http://dx.doi.org/10.1017/S0954579400006003>.
- 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)*. University of Miami, Coral Gables, FL.
- 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>.
- Mur, M., Bandettini, P.A., Kriegeskorte, N., 2009. Revealing representational content with pattern-information fMRI—an introductory guide. *Soc. Cogn. Affect. Neurosci.* 4 (1), 101–109, <http://dx.doi.org/10.1093/scan/nsn044>.
- 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 Autism Other Dev. Disabil.* 23 (1), 5–14, <http://dx.doi.org/10.1177/1088357607311443>.
- Nation, K., Penny, S., 2008. Sensitivity to eye gaze in autism: is it normal? Is it automatic? Is it social? *Dev. Psychopathol.* 20 (01), 79–97, <http://dx.doi.org/10.1017/S0954579408000047>.
- Oberwilling, E., et al., 2016. Look into my eyes: investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *Neuroimage* 130, 248–260, <http://dx.doi.org/10.1016/j.neuroimage.2016.02.026>.
- 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. *Neurosci. Biobehav. Rev.* 37 (10), 2516–2528, <http://dx.doi.org/10.1016/j.neubiorev.2013.07.017>.
- 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.
- Posner, M.I., 1980. *Orienting of attention*. *Q. J. Exp. Psychol.* 32 (1), 3–25.
- Redcay, E., Dodel-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>.

- 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. *Hum. Brain Mapp.* 34, 2511–2523, <http://dx.doi.org/10.1002/hbm.22086>.
- 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. *Front. Integr. Neurosci.* 4, 127, <http://dx.doi.org/10.3389/fnint.2010.00127>.
- 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. *J. Cogn. Neurosci.* 22 (12), 2702–2715, <http://dx.doi.org/10.1162/jocn.2009.21401>.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., Vogeley, K., 2013. Toward a second-person neuroscience. *Behav. Brain Sci.* 36 (4), 393–414, <http://dx.doi.org/10.1017/S0140525X12000660>.
- Schilbach, L., 2016. Towards a second-person neuropsychiatry. *Philos. Tran. Royal Soc. B* 371 (1686), 20150081, <http://dx.doi.org/10.1098/rstb.2015.0081>.
- 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>.
- Stone, W.L., Ousley, O.Y., Littleford, C.D., 1997. Motor imitation in young children with autism: what’s the object? *J. Abnorm. Child Psychol.* 25, 475–485, <http://dx.doi.org/10.1023/A:1022685731726>.
- Tomasello, M., 1995. Joint attention as social cognition. In: Moore, C., Dunham, P.J. (Eds.), *Joint Attention: Its Origins and Role in Development*. Lawrence Erlbaum Associates, Hillsdale.
- 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. *Front. Hum. Neurosci.* 6, 331, <http://dx.doi.org/10.3389/fnhum.2012.00331>.
- Waytz, A., Epley, N., Cacioppo, J.T., 2010. Social cognition unbound: insights into anthropomorphism and dehumanization. *Curr. Dir. Psychol. Sci.* 19, 58–62.
- 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>.
- 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>.