Detecting communicative intent in a computerised test of joint attention

Nathan Caruana Corresp., [1, 2, 3] , **Genevieve McArthur** [1, 2, 4] , **Alexandra Woolgar** [1, 2, 3] , **Jon Brock** [2, 4, 5, 6]

[1] Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia
[2] ARC Centre of Excellence in Cognition and its Disorders, Sydney, NSW, Australia
[3] Perception in Action Research Centre, Sydney, NSW, Australia
[4] Centre for Atypical Neurodevelopment, Sydney, NSW, Australia
[5] Department of Psychology, Macquarie University, Sydney, NSW, Australia
[6] Department of Cognitive Science, Macquarie University, Sydney, Australia

Corresponding Author: Nathan Caruana
Email address: nathan.caruana@mq.edu.au

The successful navigation of social interactions depends on a range of cognitive faculties – including the ability to achieve joint attention with others to share information and experiences. We investigated the influence that intention monitoring processes have on gaze-following response times during joint attention. We employed a virtual reality task in which 16 healthy adults engaged in a collaborative game with a virtual partner to locate a target in a visual array. In the *Search* task, the virtual partner was programmed to engage in non-communicative gaze shifts in search of the target, establish eye contact, and then display a communicative gaze shift to guide the participant to the target. In the *NoSearch* task, the virtual partner simply established eye contact and then made a single communicative gaze shift towards the target (i.e., there were no non-communicative gaze shifts in search of the target). Thus, only the Search task required participants to monitor their partner's communicative intent before responding to joint attention bids. We found that gaze following was significantly slower in the Search task than the NoSearch task. However, the same effect on response times was not observed when participants completed non-social control versions of the Search and NoSearch tasks, in which the avatar's gaze was replaced by arrow cues. These data demonstrate that the intention monitoring processes involved in differentiating communicative and non-communicative gaze shifts during the Search task had a measureable influence on subsequent joint attention behaviour. The empirical and methodological implications of these findings for the fields of autism and social neuroscience will be discussed.

Detecting Communicative Intent in a Computerised Test of Joint Attention

Nathan **Caruana**[1 2 3], Genevieve **McArthur**[1 2 4], Alexandra **Woolgar**[1 2 3], & Jon **Brock**[2 4 5]

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

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

[3] Perception in Action Research Centre, Sydney, Australia

[4] Centre for Atypical Neurodevelopment, MACQUARIE UNIVERSITY, Sydney, Australia

[5] Department of Psychology, MACQUARIE UNIVERSITY, Sydney, Australia

Corresponding Author:

Dr. Nathan Caruana

Email: nathan.caruana@mq.edu.au

ABSTRACT

The successful navigation of social interactions depends on a range of cognitive faculties – including the ability to achieve joint attention with others to share information and experiences. We investigated the influence that intention monitoring processes have on gaze-following response times during joint attention. We employed a virtual reality task in which 16 healthy adults engaged in a collaborative game with a virtual partner to locate a target in a visual array. In the *Search* task, the virtual partner was programmed to engage in non-communicative gaze shifts in search of the target, establish eye contact, and then display a communicative gaze shift to guide the participant to the target. In the *NoSearch* task, the virtual partner simply established eye contact and then made a single communicative gaze shift towards the target (i.e., there were no non-communicative gaze shifts in search of the target). Thus, only the Search task required participants to monitor their partner's communicative intent before responding to joint attention bids. We found that gaze following was significantly slower in the Search task than the NoSearch task. However, the same effect on response times was not observed when participants completed non-social control versions of the Search and NoSearch tasks, in which the avatar's gaze was replaced by arrow cues. These data demonstrate that the intention monitoring processes involved in differentiating communicative and non-communicative gaze shifts during the Search task had a measurable influence on subsequent joint attention behaviour. The empirical and methodological implications of these findings for the fields of autism and social neuroscience will be discussed.

Joint attention is defined as the simultaneous coordination of attention between a social partner and an object or event of interest (Bruner, 1974; 1995). It is an intentional, communicative act. In the prototypical joint attention episode, one person *initiates* joint attention (IJA) by pointing, turning their head, or shifting their eye gaze to intentionally guide their social partner to an object or event in the environment. The partner must recognise the intentional nature of this initiating behaviour and *respond* to that joint attention bid (RJA) by directing their attention to the cued location (Bruinsma, Koegel, & Koegel, 2004).

The ability to engage in joint attention is considered critical for the normal development of language and for navigating social interactions (Adamson, Bakeman, Deckner, & Romski, 2009; Charman, 2003; Dawson et al., 2004; Mundy, Sigman, & Kasari, 1990; Murray et al., 2008; Tomasello, 1995) and its developmental delay is a hallmark of autism spectrum disorders (Lord et al., 2000; Stone, Ousley, & Littleford, 1997). Yet despite its importance to both typical and atypical development, very little is known about the neurocognitive mechanisms of joint attention. By definition, joint attention involves an interaction between two individuals. The challenge for researchers, therefore, has been to develop paradigms that achieve the ecological validity of a dynamic, interactive, social experience, whilst at the same time maintaining experimental control.

In a recent functional magnetic resonance imaging (fMRI) study, Schilbach et al. (2010) investigated the neural correlates of joint attention using a novel virtual reality paradigm. During the scan, participants' eye-movements were recorded as they interacted with an anthropomorphic avatar. They were told that the avatar's gaze was controlled by a confederate outside the scanner also using an eye-tracking device. In fact, the avatar was controlled by a computer algorithm that responded to the participant's own eye-movements. On RJA trials (referred to as OTHER_JA by Schilbach et al., 2010), the avatar looked towards one of three squares positioned around his face, and participants were instructed to respond by looking at the same square. Participants also completed IJA trials in which the roles were reversed.

Similar tasks have been used in other fMRI studies using either gaze-contingent avatars (Oberwelland et al., 2016) or live-video links to a real social partner (Redcay et al., 2012; Saito et al., 2010). Together, these interactive paradigms represent an important step towards an ecologically valid measure of joint attention. There is, however, a potentially important limitation of the tasks used in these studies: in each task, every trial involved a single unambiguously communicative eye-gaze cue. On RJA trials, the participant's partner would make a single eye-movement towards the target location and the participant knew they were required to respond to that isolated cue. This differs from real-life joint attention episodes, which are embedded within complex ongoing social interactions. In real life, responding to a joint attention bid requires that the individual first identifies the intentional nature of their partner's behaviour. That is, they must decide whether or not the cue is one they should follow. We refer to this component of joint attention as "intention monitoring".

In a recent fMRI study, we developed a novel joint attention task to better capture this intention monitoring process (Caruana, Brock & Woolgar, 2015).

Following Schilbach et al. (2010), participants played a cooperative game with an avatar whom they believed to be controlled by a real person (referred to as “Alan”), but was actually controlled by a gaze-contingent algorithm. The participant and avatar were both allotted onscreen houses to search for a burglar (see Figure 1). On IJA trials, the participant found the burglar, made eye contact with the avatar, and then guided the avatar to the burglar by looking back at the house in which the burglar was hiding. On RJA trials, the participant found all of their allotted houses to be empty. Once Alan had finished searching his own houses, he would make eye contact with the participant before guiding them towards the house containing the burglar.

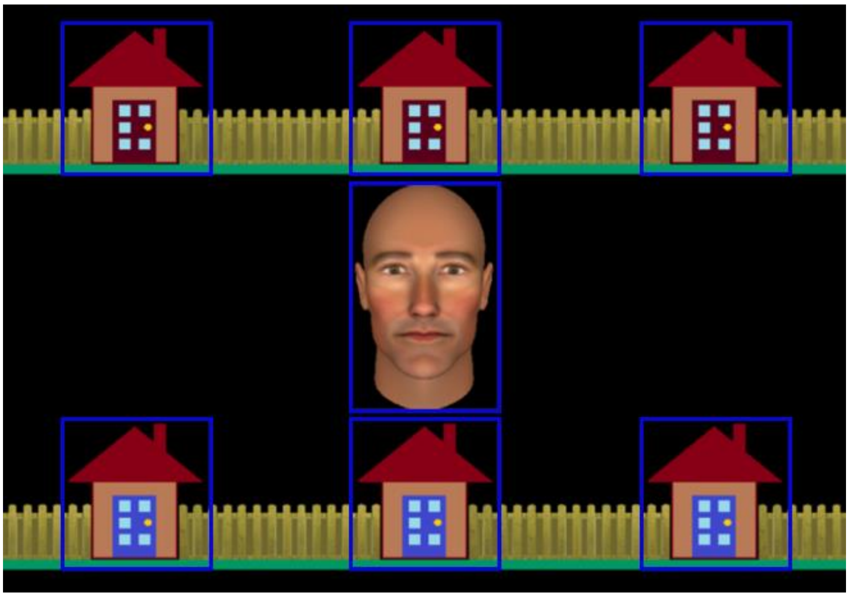

Figure 1. Experimental display showing the central avatar (“Alan”) and the six houses in which the burglar could be hiding. Gaze areas of interest (GAOIs), are represented by blue rectangles, and were not visible to participants.

The critical innovation of this task was the initial search phase. This provided a natural and intuitive context in which participants could determine, on each trial, their role as either the responder or initiator in a joint attention episode (previous studies had provided explicit instructions; e.g., Schilbach et al., 2010; Redcay et al., 2012). More importantly for current purposes, the RJA trials required participants to monitor their partner's communicative intentions. During each trial, the avatar made multiple non-communicative eye-movements as he searched his own houses. The participant had to ignore these eye-movements and respond only to the communicative "initiating saccades" that followed the establishment of eye contact. This is consistent with genuine social interactions in which eye contact is used to signal one's readiness and intention to communicate, particularly in the absence of verbal cues (Cary, 1978).

We compared the RJA trials in this new paradigm to non-social control trials (referred to as RJAc) in which the eye gaze cue was replaced by a green arrow superimposed over the avatar's face. Analysis of saccadic reactions times revealed that participants were significantly slower (by approximately 209 ms) to respond to the avatar's eye gaze cue than they were to respond to the arrow (Caruana et al., 2015). We have since replicated this finding in an independent sample of adults and, intriguingly, found that the effect is exaggerated in a group of autistic individuals (Caruana, Ham et al., submitted).

One explanation for these findings is that they reflect the intention monitoring aspects of RJA. Specifically, participants are slower to respond to eye gaze cues than arrows because it takes time to identify the cue as being an intentional and communicative bid to initiate joint attention. In the control condition, the arrow presents an unambiguous attention cue, and so the participant does not need to decide whether they should respond to it or not. The implication here is that intention monitoring is a cognitively demanding operation that requires time to complete and is manifest in the response times to eye gaze cues. However, before reaching such a conclusion, it is important to rule out alternative explanations. In particular, it may be that responses in the control condition were relatively fast simply because the arrow cue was more salient than the corresponding gaze cue. In the current study, therefore, we aimed to pit the intention monitoring and perceptual salience accounts against one another.

To this end, we tested a new sample of participants using the same task but with one further manipulation. On half the trials, we eliminated the search phase of the task. Thus, on RJA trials, the avatar only made a single eye movement to the target to initiate joint attention, and participants knew unambiguously that they should follow it. As such, this 'NoSearch' version of the task closely resembled previous joint attention paradigms that have not attempted to capture intention monitoring processes (e.g., Schilbach et al., 2010; Redcay et al., 2012).

If the slower responses to eye gaze in our previous studies was entirely a function of the perceptual salience of the two cues, then this manipulation should have no effect on saccadic reaction times. However, if intention monitoring processes play a role, removing the search phase should significantly reduce response times for the RJA condition, but should have no effect on response times for the RJAc condition because the arrow cue is unambiguous whether or not it is preceded by a search phase. In other words, a condition (Social vs. Control) by task (Search vs. NoSearch) interaction would indicate that response times to joint attention bids are influenced by the intention monitoring processes that pre-empt RJA behaviours.

Method

Ethical Statement

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

Participants

Sixteen right-handed adults with typical development, normal vision, and no history of neurological impairment participated in this study (3 female, M_{age} = 19.92, *SD* = 1.03).

Stimuli

We employed an interactive paradigm that we had previously used to investigate the neural correlates of RJA and IJA (Caruana, et al., 2015). The stimuli comprised an anthropomorphic avatar face, generated using *FaceGen* (Singular Inversions, 2008), that subtended 6.5 degrees of visual angle. The avatar's gaze was manipulated so that it 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 were arranged in two horizontal rows above and below the avatar and each subtended four degrees of visual angle.

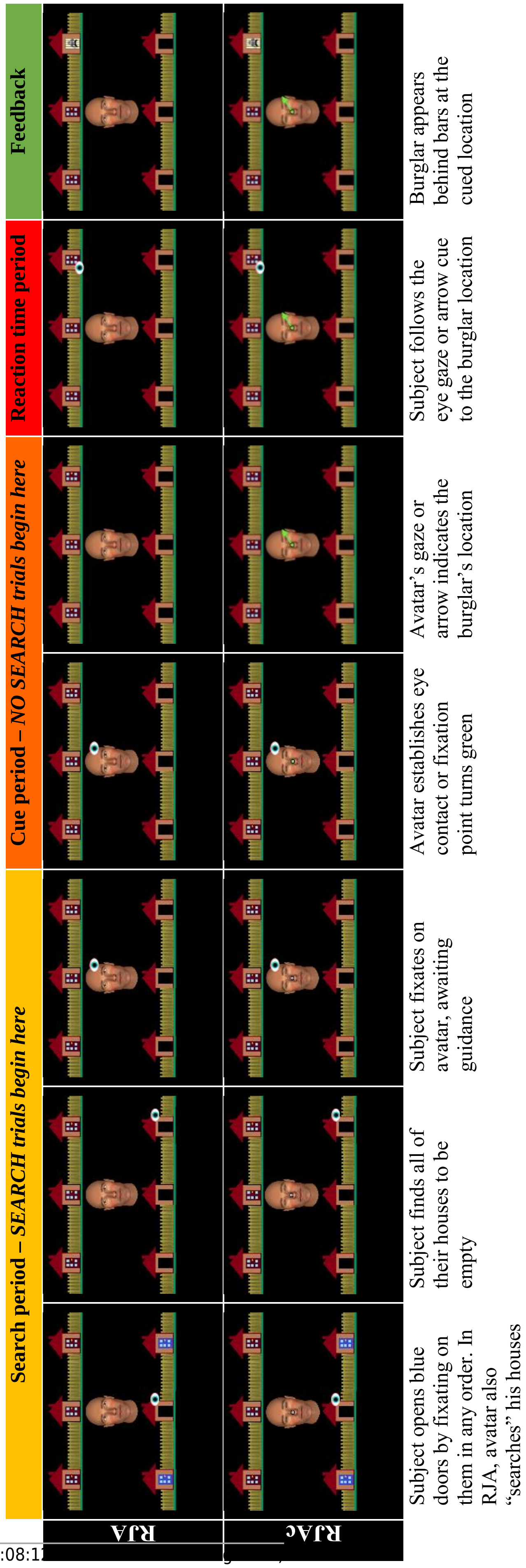

Figure 2. Event timeline for RJA and RJAc trials.

Joint Attention Task

Social Conditions (RJA and IJA). Participants played a cooperative “Catch-the-Burglar” game with an avatar whom they believed was controlled by another person named “Alan” in a nearby eye tracking laboratory using live infrared eye tracking. In reality, a gaze-contingent algorithm controlled the avatar’s responsive behaviour (see Caruana et al., 2015 for a detailed description of this algorithm and a video demonstration of the task). The goal of the game was to catch a burglar that was hiding inside one of the six houses presented on the screen. Participants completed two versions of the social conditions (i.e., Search and NoSearch tasks) during separate blocks.

Search Task. This task was identical to the “Catch-the-Burglar” task employed in our previous work (e.g., Caruana et al., 2015). Each trial in the Search task began with a “search phase”. During this period, participants saw two rows of houses on a computer screen including a row of three blue doors and a row of three red doors. They were instructed to search the row of houses with blue doors while Alan searched the row of houses with red doors. Participants were told that they could not see the contents of Alan’s houses and that Alan could not see the contents of their houses. Whoever found the burglar first had to guide the other person to the correct location.

Participants searched the houses with blue doors in any order by fixating on them. Once a fixation was detected on a blue door, it opened to reveal either the burglar or an empty house. On some trials, only one or two blue doors were visible, whilst the remaining doors were already open. This introduced some variability in the order with which participants searched their houses that made Alan's random search behaviour appear realistically unpredictable.

Once the participant fixated back on the avatar's face, Alan was programmed to search 0-2 more houses and then make eye contact. This provided an interval in which participants could observe Alan's non-communicative gaze behaviour as he completed his search. The onset latency of each eye movement made by Alan was jittered with a uniform distribution between 500-1000 ms.

On RJA trials, participants discovered that all of their allotted houses were empty (Figure 2, row 1), indicating 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 in random order before establishing eye contact with the participant. Alan then initiated joint attention by directing his gaze towards one of his allotted houses. If the participant made a "responding saccade" and fixated the correct location, the burglar was captured.

On IJA trials, the participant found a burglar behind one of the blue doors. They were then required to fixate back on Alan's face, at which point the door would close again to conceal the burglar. Again, Alan was programmed to search 0-2 houses before looking straight at the participant. Once eye contact was established, participants could initiate joint attention by making an "initiating saccade" to fixate on the blue door that concealed the burglar. Alan was programmed to only respond to initiating saccades that followed the establishment of eye contact, and to follow the participant's gaze, irrespective of whether the participant fixated the correct house or not. Whilst performance on IJA trials was not of interest in the current study, the inclusion of this condition created a context for the collaborative search element of the task and allowed direct comparison with our previous studies in which participants alternated between initiating and responding roles.

When the participant made a responding or initiating saccade to the correct location, the burglar appeared behind prison bars to indicate that he had been successfully captured (e.g., Figure 2, column 7). However, the burglar appeared in red at his true location, to indicate that he had escaped, if participants (1) made a responding or initiating saccade to an incorrect location, (2) took longer than three seconds to make a responding or initiating saccade, or (3) spent more than three seconds looking away from task-relevant stimuli (i.e., Alan and houses). Furthermore, trials were terminated if the participants took longer than three seconds to begin searching their houses at the beginning of the trial. On these trials, red text reading "Failed Search" appeared on the screen to provide feedback.

NoSearch Task. This version of the task was identical to the Search task except that the search phase in each trial was removed. In IJA trials, all but one house was visibly empty (i.e., the door was open and no burglar was present), and participants were instructed that if they saw a blue door in their allotted row of houses that the burglar would be "hiding" behind it. In RJA trials, all of the houses were visibly empty. For both IJA and RJA trials, Alan's eyes would be closed at the beginning of the trial, and then open after 500-1000 ms (jittered with a uniform distribution) so that he was looking at the participant. Alan would then wait to be guided on IJA trials. On RJA trials, Alan shifted his gaze to guide the participant after a further 500-1000 ms, provided that eye contact had been maintained. Thus, in both the Search and NoSearch tasks, Alan made eye contact with the participant before guiding them to the burglar on RJA trials. Therefore, the perceptual properties of the gaze cue itself were identical between tasks, but the NoSearch task removed the requirement to use the eye contact cue to identify communicative gaze shifts.

Control Conditions (RJAc and IJAc). For each of the social conditions in both versions of the task, we employed a control condition that was closely matched on non-social task demands (e.g., attentional orienting, oculomotor control). In these conditions (RJAc and IJAc), participants were told that they would play the game without Alan, whose eyes remained closed during the trial. In the Search task, a grey fixation point was presented over the avatar's nose until the participant completed their search and fixated upon it. After a short delay, the fixation point turned green (analogous to the avatar making eye contact). From this point onwards, the Search and NoSearch tasks were identical. On IJAc trials, the green fixation point was the cue to saccade towards the burglar location. On RJAc trials a green arrow subtending three degrees of visual angle cued the burglar's location (analogous to Alan's guiding gaze; see Caruana et al., 2015 for a video with example trials from each condition).

Procedure

Joint attention task. The experiment was presented using *Experiment Builder* 1.10.165 (SR Research, 2004). Participants completed four blocks, each comprising

108 trials: two blocks involved the Search task, and another two blocks involved the NoSearch task. Search and NoSearch block pairs were presented consecutively, however their order was counterbalanced across participants. Within each pair of Search and NoSearch blocks, one block required the participant to monitor the upper row of houses, and the other required them to monitor the lower row of houses.

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. Each cluster began with a cue lasting 1000 ms that was presented over the avatar stimulus and read “Together” for a social cluster and “Alone” for a control cluster. Trial order randomisation was constrained to ensure that the location of the burglar, the location of blue doors, and the number of gaze shifts made by the avatar were matched within each block and condition.

Eye tracking. Eye-movements from the right eye only were recorded with a sampling rate of 500Hz using a desktop-mounted EyeLink 1000 Remote Eye-Tracking System (SR Research Ltd., Ontario, Canada). Head movements were stabilised using a chinrest. We conducted an eye tracking calibration using a 9-point sequence at the beginning of each block. Seven areas of interest (AOIs) over the houses and avatar stimulus were used by our gaze-contingent algorithm (see Caruana et al., 2015 for details). A recalibration was conducted if the participant made consecutive fixations on the borders or outside the AOIs. Trials requiring a recalibration were excluded from all analyses.

Subjective task ratings. After playing the interactive game, participants rated the difficulty, naturalness, intuitiveness, and pleasantness of the social and control conditions on a 5-point Likert scale (1 = not at all, 5 = extremely). They also rated whether they preferred playing the game with Alan or alone on a 10-point scale (0 = complete preference for social interaction, 10 = complete preference for completing the task alone). Participants were also asked to rate how co-operative Alan was on the same scale. This provided an opportunity for participants to declare whether they disbelieved that Alan was a real person. At the end of the study, participants were told that Alan was controlled by a computer program. They were then asked to rate how convinced they were that Alan had been controlled by another human on a 5-point scale (one rating for the whole experiment).

Scores

Accuracy. We calculated the proportion of trials where the participant succeeded in catching the burglar in each condition (i.e., RJA and RJAc) for each task separately (i.e., Search and NoSearch). We excluded from the accuracy analysis any trials that required a recalibration or (in the Search task) were failed due to an error during the search phase.

Saccadic Reaction Times. For correct trials, we measured the latency (in ms) between the presentation of the gaze cue (for RJA trials) or the arrow cue (for RJAc trails), and the onset of the participant's responding saccade towards the burglar location (see Figure 2, Reaction time period).

Statistical Analyses

Repeated measures Analysis of Variance (ANOVA) tests were used for all measures of task performance using the ezANOVA (ez) package in R (Lawrence, 2013), reporting the generalised eta squared (η_G^2) measure of effect size. Significant task*condition interactions effects were followed-up with Welch's two sample unequal variances t-tests (Welch, 1947) that assessed the effect of condition (i.e., RJA versus RJAc) for each task (i.e., Search versus NoSearch). For saccadic reaction time measures we report analyses of the mean reaction time, having excluded trials with dwell times less than 150 ms (trials timed out after 3000 ms). Full syntax and output for these analyses can be found in the Rmarkdown document (Supplementary Material 1).

Non-parametric tests (related samples Wilcoxon Signed Ranks Test) were used on all measures of subjective experience since these data violated assumptions of normality. A significance criterion of $p < 0.05$ was used for all analyses.

Results

Summary statistics for all dependant variables are shown in Table 1. For each analysis, we report the main effect of task (Search vs. NoSearch), the main effect of condition (RJA vs. RJAc), and task*condition interaction.

Table 1.

Means and standard deviations for outcome measures by condition and task.

	RJA		**RJAc**	
	Search	**NoSearch**	**Search**	**NoSearch**
Accuracy (%)	93.17 (6.68)	97.57 (2.31)	94.79 (6.53)	98.15 (3.02)
Saccadic Reaction Time (ms)	585.57 (133.10)	457.46 (80.62)	336.62 (40.32)	349.20 (45.11)

Note. Accuracy is reported as % of trials that were correct. Saccadic reaction times are reported in milliseconds. Summary statistics are provided in the format *M(SD)*.

Accuracy. As depicted in Figure 3A, participants made significantly more errors on the Search task than the NoSearch task (main effect of task, $F(1,15) = 11.81$, $p = .004$, $\eta_G^2 = 0.14$). There was no significant main effect of condition ($F(1,15) = 2.99$, $p = .104$, $\eta_G^2 = 0.03$), nor a condition*task interaction ($F(1,15) = 0.59$, $p = .453$, $\eta_p^2 = 0.01$). Therefore participants were less accurate in the Search task, and there was no evidence that this differed between social and control conditions.

Saccadic reaction time. As depicted in Figure 3B, participants were significantly slower on the Search task than the NoSearch task (main effect of task (F(1,15) = 11.07, p = .005, η^2_G = 0.13). They were also significantly slower to respond on RJA trials than RJAc trials overall (main effect of condition, F(1,15) = 98.75, p < .001, η^2_G = 0.57). Importantly, there was also a significant task*condition interaction (F(1,15) = 43.86, p < .001, η^2_G = 0.18), indicating a larger effect of task in the social condition. Follow-up paired t-tests revealed that responses to social gaze were significantly slower in the Search task than the NoSearch task (t(15) = 4.82, p < .001), whereas response times to arrow cues did not significantly differ between the two versions of the control task (t(15) = -0.85, p = .411). Nonetheless, there was still an effect of condition (i.e., slower responses to gaze cues than arrow cues) in NoSearch (t(15) = 8.51, p < .001) as well as Search (t(15) = 9.31, p < .001) tasks.

341

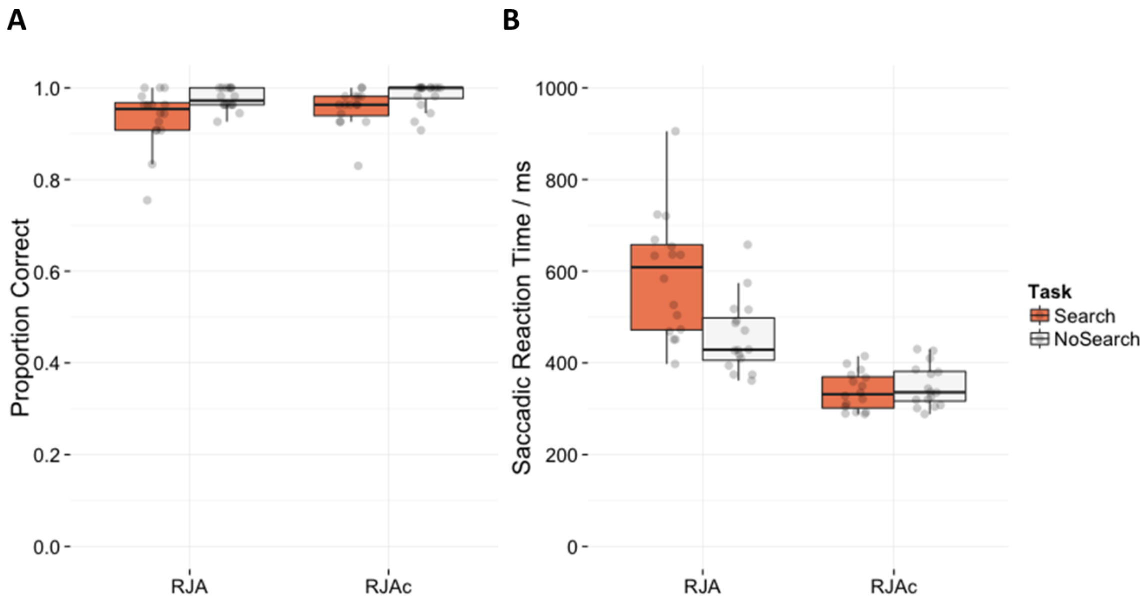

Figure 3. Box plots displaying (A) average proportion of correct trials, and (B) average saccadic reaction times in RJA and RJAc conditions, separated by task (i.e., Search, NoSearch). Data points represent individual participant means.

342

Experience ratings. For both the social and control conditions, there were no significant differences between the Search and NoSearch tasks on ratings of task difficulty, naturalness, intuitiveness, pleasantness, or cooperativeness (all *p*s > .054). Summary statistics are reported in Table 2.

Table 2

Summary statistics for subjective task ratings by condition and task.

	Task Version		Statistics	
	Search	**NoSearch**	***V***	***p***
Social Condition				
Difficulty	2.0	1.0	38	.066
Naturalness	4.0	4.0	16	.777
Intuitiveness	4.0	4.5	16	.243
Pleasantness	4.0	3.0	12	.824
Cooperativeness	5.0	5.0	9	.766
Control Condition				
Difficulty	1.0	1.0	18.5	.106
Naturalness	4.0	4.0	20	.803
Intuitiveness	4.5	5.0	3	.066
Pleasantness	3.0	3.5	10.5	.588

Note. Experience ratings were provided on a 5-point scale (1 = not at all, 5 = extremely). Task effects were tested using a two-sample Wilcoxon signed rank test with continuity correction.

Belief that the avatar was controlled by a real person. All participants said that they were convinced that the avatar was controlled by a real person, and rated the strength of this belief as either a 4/5 or 5/5 (*median = 5*). The five participants who provided a 4/5 rating, commented that they considered the possibility that Alan could be computer-controlled, but disregarded it. Participants' beliefs about whether they are engaged in a genuine social interaction may be important, and have been found to influence the neural processing of an avatar's gaze behaviour in the context of joint attention interactions (Caruana, de Lissa et al., 2016).

Preference for social condition over control condition. On average, participants reported preferences for playing the interactive task (i.e., social condition), rather than on their own (i.e., control condition). This preference was statistically equivalent for both the Search (*median = 3*) and NoSearch tasks (*median = 3.5*; $V = 40.5$, $p = .750$).

Discussion

One of the main challenges facing social neuroscience – and the investigation of joint attention in particular – is the need to achieve ecological validity whilst maintaining experimental control. During genuine joint attention experiences, our social cognitive faculties are engaged whilst we are immersed in complex interactions consisting of multiple social cues with the potential for communication. A critical but neglected aspect of joint attention is the requirement to identify those cues that are intended to be communicative. In the specific case of eye gaze cues, the responder must differentiate gaze shifts that signal an intentional joint attention bid from other, non-communicative gaze shifts. The results of the current study indicate that this intention monitoring process has a measurable effect on responsive joint attention behaviour.

The Search version of our Catch-the Burglar task was identical to that used in our previous studies. In the social (RJA) condition, participants found all of their houses to be empty, waited for their partner, Alan, to complete his search, make eye contact, and then guide them to the burglar's location. We replicated our previous finding (Caruana et al., 2015) that participants were slower to respond in this condition than in the matched control (RJAc) condition in which the avatar's eye gaze cue was replaced by an arrow.

The addition of the NoSearch version of the task allowed us to differentiate between two competing explanations for this effect. The difference in response times for eye gaze (RJA) versus arrow (RJAc) cues remained, even when the search phase was removed from the task. This suggests that the arrow cue was indeed a more perceptually salient cue. Importantly, however, the magnitude of the effect was significantly reduced, as indicated by the task by condition interaction. This interaction cannot be explained in terms of perceptual salience because the eye gaze and arrow cues were identical in the Search and NoSearch tasks.

The interaction can also be viewed by contrasting the effect of task (Search vs. NoSearch) for the two different conditions (RJA or RJAc). In the RJA condition, participants were significantly faster to respond to the eye gaze cue when the search phase was removed. This is consistent with our intention monitoring account. The search phase required the participant and their virtual partner to make multiple non-communicative eye-movements prior to the joint attention episode. Participants therefore had to differentiate between eye-movements made by the avatar that signalled a communicative joint attention bid and those that were merely a continuation of their search. In the NoSearch task, every eye-movement made by the avatar was communicative, thereby removing the requirement to monitor his communicative intent, enabling faster response times.

Importantly, there was no effect of task (Search vs NoSearch) for the RJAc condition. This allows us to discount a number of alternative explanations for the task effect in the RJA condition. For example, it could be argued that the slower responses in the Search task reflected differences in the timing of the stimulus presentation (e.g., the delay between participants fixating on the avatar and the avatar making his guiding saccade). However, the timing of the stimuli were programmed to be identical in the corresponding RJA and RJAc conditions of the Search and NoSearch tasks, so any effect of stimulus timing should have been evident in both conditions. Another plausible explanation is that participants were slower in the Search task because this required them to switch from searching for the burglar to responding to the avatar on each trial. But again this applied equally to the RJA and RJAc conditions, so it cannot explain the task by condition interaction.

In short, the observed interaction between task and condition is entirely consistent with our intention monitoring account and cannot be explained in terms of the perceptual salience of different cues, the timing of the stimulus presentation, or the requirement to switch between searching and responding.

The current data also help to clarify some of our other recent findings in studies using the Search version of our interactive task. In one study employing the

Search version of our task, we used fMRI to investigate the neural correlates of RJA (Caruana et al., 2015). By contrasting activation in the RJA (eye gaze) and RJAc (arrow) conditions, we identified a broad frontotemporoparietal network including the right temporo-parietal junction and right inferior frontal lobe. These brain regions are strongly associated with aspects of social cognition including mentalising (e.g., Saxe & Kanwisher, 2003) and predicting anothers' actions (Danckert et al., 2002; Hamilton & Grafton, 2008) but have not been previously linked to RJA (cf. Redcay et al., 2012; Schilbach et al., 2010). The tasks used in previous studies of RJA were similar to the current NoSearch task. As such, they would not have captured the intention monitoring processes involved in RJA, perhaps explaining the discrepenacy with our fMRI study. Future neuroimaging studies of joint attention could employ the current study's design, and compare activation observed during the Search and NoSearch task. If our interpretation is correct then removing the search component should reduce the involvement of temporoparietal and inferior frontal regions in the RJA condition.

In another study (Caruana, Ham, et al., forthcoming), we investigated joint attention in adults with autism. Observational studies of real-life interactions provide overwhelming evidence that joint attention impairments are a core feature of autism (Charman et al., 1997; Dawson et al., 2004; Loveland & Landry, 1986; Mundy et al., 1990; Osterling, Dawson, & Munson, 2002; Wong & Kasari, 2012). However, previous computer-based experimental studies of joint attention have largely failed to find consistent evidence of gaze following difficulties (Leekam, 2015; Nation &

Penny, 2008). One possible explanation for this is that autistic individuals have an underlying difficulty in understanding the social significance or communicative intentions conveyed by eye contact (cf. Böckler, Timmermans, Sebanz, Vogeley, & Schilbach, 2014; Senju & Johnson, 2009) that is not captured by the tasks used in previous studies of autism. In contrast to these studies, we did find evidence of impairment: autistic adults made more errors and were slower to respond than control participants in the RJA condition despite showing no impairment on the control condition. Future studies involving individuals with autism and the Search and NoSearch versions of our task would clarify this issue further.

Summary

In everyday joint attention episodes, a critical aspect of responding to joint attention bids is the ability to discern which social cues have communicative intent and which do not. The results of the current study indicate that this intention monitoring component has a measureable effect on responding behaviour. Moreover, this component can be isolated by contrasting joint attention episodes occurring in the context of a realistically complex social interaction versus a simplified context in which each cue is unambiguously communicative. The clear differences in performance on the Search and NoSearch versions of our task highlight the importance of striving for ecological validity in studies of social cognition (cf. Schilbach et al., 2013). The results also demonstrate the potential of our task for investigating the different components of joint attention in typically developing children and in clinical populations associated with atypical social cognition.

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 and Developmental Disorders, 39*(1), 84-96. doi: 10.1007/s10803-008-0601-7

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

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

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.

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. (2016). Beliefs about human agency influence the neural processing of gaze during joint attention. *Social Neuroscience,* doi:10.1080/17470919.2016.1160953

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

Danckert J, Ferber S, Doherty T, Steinmetz H, Nicolle D, Goodale MA (2002) Selective, non-lateralized impairment of motor imagery following right parietal damage. *Neurocase 8*, 194–204

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

Hamilton AFC, de, Grafton ST. 2008. Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex*, *18* (5), 1160–1168.

Lawrence MA (2013). ez: Easy analysis and visualization of factorial experiments. R package version 4.2-2, URL http://CRAN.R-project.org/package=ez.

Leekam, S. (2015). Social cognitive impairment and autism: what are we trying to explain? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *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. *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 and 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

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

Oberwelland, E., et al. (2016). Look into my eyes: Investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *NeuroImage*, in press.

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

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.

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

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.

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.

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.

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

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

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

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.

Welch, B. L. (1947). The generalization of ‘student's’ problem when several different population variances are involved. *Biometrika, 34*, 28-35.

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