Evaluating the relevance of gaze cues in context: A study of virtual joint attention interactions in autism and typical development

Ayeh Alhasan, BSc. Chemical Engineering

Supervised by: Dr. Nathan Caruana

Co-supervisor: Prof. Liz Pellicano

Department of Cognitive Science

Faculty of Human Sciences

Macquarie University, Sydney, Australia

This dissertation is presented for the degree of Master of Research (MRes)

November, 2019

Table of Contents

Abstract

Joint attention – the ability to attend to the same thing as others – is an important social ability and its delayed development is characteristic in autism. This may reflect a specific difficulty in identifying relevant communicative cues (e.g., eye movements) embedded in a realistic social interaction. This project investigated this hypothesis by utilising an ecologically valid joint attention paradigm. Participants played a co-operative interactive game with an on-screen avatar which required the participant to evaluate and respond to the eye gaze behaviour of their partner. In Experiment1, neurotypical adults played three different versions of this game manipulating the contextual clues available prior to the joint attention bid. Here, the context conditions included non-communicative eye movements which were either (1) informative, by being predictive of the target's location (Predictive Search), (2) non-informative, not predictive of the target's location (Random Search), or (3) did not contain non-communicative eye movements before the joint attention bid (NoSearch). Each context was performed once with each stimulus (Eyes and Arrows). Data was analysed for accuracy and saccadic reaction times (SRT) in response to joint attention bids. Results revealed that, overall, participants made more errors with the Random Search context than both the NoSearch and Predictive contexts. They were also significantly faster to respond on the Predictive Search than the NoSearch and Random Search contexts with both Eyes and Arrows stimuli. Critically, the disadvantage (i.e., slower reaction time) for embedding Eyes in Random context was smaller than Arrows compared to NoSearch baseline. Additionally, the advantage (i.e., faster reaction time) for embedding Eyes in the Predictive context was larger than Arrows compared to the NoSearch baseline. These findings collectively suggest a relative advantage for identifying relevant Eyes rather than Arrows when embedded in a realistic context. This relative advantage could be attributed to the unique advantage of eye contact as an ostensive signal.

In Experiment 2, we asked young autistic people to play the same task, except that we investigated two contexts only (i.e., Random and NoSearch) and two stimuli (i.e., Eyes, Arrows). Comparisons between responsivity of the autistic group to a neurotypical comparison group revealed no overall significant group differences in terms of accuracy and SRT. Participants in both groups made more errors with the Random compared to NoSearch context. They were also slower to respond on the Random Search compared to NoSearch. Investigating the trend of effects for SRT in each group separately revealed that the neurotypical group showed the same relative advantage for Eyes compared to Arrows as in Experiment 1. This was again characterised by a smaller effect of Random context on responsivity for Eyes than Arrows. This relative advantage, however, was not replicated in the autistic group. This finding reveals that although young autistic individuals were able to complete the tasks with performance that is comparable to their neurotypical peers, they seem to lack sensitivity to eye contact as an ostensive signal. These findings are critical in understanding the specific factors that contribute to the difficulty faced by autistic people in responding to joint attention. More work is needed to verify these findings using larger sample sizes. Future studies should also investigate the influence of a Predictive context (as in Experiment 1) on joint attention to further understand how informative contextual clues affect responsivity in autism.

Statement of Originality

I, Ayeh Alhasan, certify that this thesis titled "Evaluating the relevance of gaze cues in context: A study of virtual joint attention interactions in autism and typical development" has not previously been submitted for a higher degree or as part of requirements for a degree to any university or institution other than Macquarie University. I also certify that this thesis is an original piece of research that has been written by me and contains no material previously published or written by another person except where due reference is made in the thesis itself. All the help and assistance that I have received in the preparation of this thesis has been properly acknowledged. This project was approved by the Macquarie University Human Ethics Committee, ID: **3775**

Ayeh Alhasan

November, 2019

Acknowledgements

To my supervisor, Dr. Nathan Caruana, for dedicating your time to support me throughout this journey. Your guidance has been paramount to the successful completion of this project. You have gone above and beyond to provide me with the support I needed and I truly appreciate it. To my co-supervisor, Professor Liz Pellicano, for sharing your expertise and critical thought to best guide and support this project. To Peter Humburg, for providing much needed support on handling the data analysis. To Marcus Ockenden, for providing technical support and never failing to save the day on numerous occasions. To Lesley McKnight, for helping sort out the nitty-gritty of administrative requirements. To Nicola, Kirilee, Ioanna, Dave and Catherine for countless precious memories and the friendship you offered throughout the past two years.

And finally, to my number one support system, my family. To my parents, Yaser and Mary, my role models. For providing me with everything I need to achieve anything I want. For being there whenever and wherever, to guide, counsel and support. For always believing in me and challenging me to achieve my best. For showing me what truly matters in life, and what doesn't. For you I am forever grateful. To my husband, Hammam, my backbone and safe spot. For helping me dream big and supporting my dreams no matter how unrealistic they may seem. For bringing me joy when I needed it the most and for encouraging me every step of the way. To my brothers, Ammar and Omar, for listening to me rant and helping me think outside the box. For always offering to help in any and every way possible.

Chapter 1: Introduction

Joint attention refers to our ability to intentionally coordinate our attention with a social partner so that we are both attending to the same object or event at the same time (Bruner, 1974; Carpenter & Liebal, 2011; Hobson, 2005; Tomasello, 1995). Typically, gaze-based joint attention involves one person initiating joint attention (IJA) by producing a gaze shift to guide a second person to an object or event of interest. The second person then responds to the joint attention bid (RJA) by attending to the same object or event (Bruinsma, Koegel, & Koegel, 2004). The early development of joint attention is pivotal in supporting the later development of language (Akhtar, Dunham, & Dunham, 1991; Charman, 2003; Dawson et al., 2004) and social-cognitive skills, including the ability to represent the mental states and perspectives of others (Mundy, 2003, 2016, 2018; Mundy & Jarrold, 2010; Mundy & Neal, 2000; Mundy & Newell, 2007). Further, since joint attention supports the ability to share experiences with others, it is critical in supporting social and cultural learning (Csibra & Gergely, 2009; Tomasello & Carpenter, 2007; Tomasello, Kruger, & Ratner, 1993). The importance of joint attention further extends into adulthood by supporting social cognitive processes which depend on social perspective-taking and coordination (e.g. Corbetta, Patel, & Shulman, 2008; Krall et al., 2015; Spreng, Mar, & Kim, 2009; Tomasello & Carpenter, 2007).

Although joint attention may be established using several different social cues (e.g., pointing or verbal cues), the first cue to be used during early joint attention interactions is eye gaze (Mundy & Jarrold, 2010). Indeed, newborn babies (D'Entremont, Hains, & Muir, 1997) and infants (Farroni, Massaccesi, Pividori, & Johnson, 2004) demonstrate the ability to follow eye gaze. However, the more advanced ability to understand and coordinate attention with others using eye gaze develops later in infancy and supports our ability to identify joint attention targets signalled by others (Heal, 2005). Autistic children are characterised by a reduced propensity to make and regulate eye contact (Adrien et al., 1993; Kanner, 1943; Mirenda, Donnellan, & Yoder, 1983; Zwaigenbaum et al., 2005), and a delay in the development of joint attention ability (e.g. Charman, 2003; Hobson & Hobson, 2007; Mundy, Sigman, & Kasari, 1994). These delays have been linked to downstream delays in the development of social cognition and language (e.g. Delinicolas & Young, 2007; Kwisthout, Vogt, Haselager, & Dijkstra, 2008). Hence, delayed joint attention is considered one of the earliest and most reliable markers of autism (Lord et al., 2000; Osterling & Dawson, 1994).

Delays in the development of joint attention responsivity in autism have also been suggested to contribute to delays in cognitive development more broadly due to reduced opportunities to learn experientially through others (Frith & Frith, 2010; Mundy & Neal, 2000; Sodian, Schuwerk, & Kristen, 2015; Vivanti, Fanning, Hocking, Sievers, & Dissanayake, 2017). For instance, there is compelling evidence indicating that early emergence of RJA behaviour, but not IJA, is a reliable predictor of later vocabulary development (Adamson, Bakeman, Suma, & Robins, 2019; Bottema-Beutel, 2016; Murray et al., 2008). Therefore, understanding the cognitive mechanisms which explain reduced responsivity to gaze-cued joint attention is fundamental for formulating empirically-informed learning paradigms to address these social communication challenges in autism. While most studies of joint attention in autism have set out to investigate this challenge, there have been inconsistent findings reported by studies using different methodological approaches (see Birmingham, Ristic, & Kingstone, 2012; Leekam, 2016; Nation & Penny, 2008 for reviews). This has resulted in little progress being made in understanding the specific factors that contribute to the delayed development of this ability.

Most joint attention studies in autism have used observational methods of natural and semi-structured social interactions. These studies have consistently reported diminished responsivity to joint attention bids in young autistic children (Charman et al., 1997; Clifford

& Dissanayake, 2008; Dawson et al., 2004; Loveland & Landry, 1986; Mundy, Sigman, & Kasari, 1990; Osterling & Dawson, 1994; Osterling, Dawson, & Munson, 2002; Wong & Kasari, 2012). For instance, Clifford and Dissanayake (2008) investigated the early development of joint attention, eye contact and affect during the first two years of development in infants later diagnosed with autism. Home videos were analysed and parental interviews conducted. Atypical gaze and affect behaviour (e.g., initiating and responsive smiles) were reported in autistic infants during the first six months, which increased in severity as they developed. Difficulties in joint attention were also reported into the second year of life. The use of such naturalistic observational paradigms in studies such as this one are highly ecologically valid, but also lack experimental control. This is problematic as it makes it challenging to manipulate and test specific aspects of a social interaction separately. This, in turn, makes it challenging to measure more objectively the consequence of the tested manipulations. Also, maintaining the consistency of the presentation of stimuli across trials and participants in these studies cannot be guaranteed.

Other studies have utilised structured experimental designs to achieve a controlled investigation of gaze responsivity in autistic people. One of the most commonly-used experimental paradigms is the gaze-cueing paradigm (e.g. Friesen & Kingstone, 1998; Langton & Bruce, 1999). This task is based on the Posner-cueing paradigm (Posner, 1980) in which participants are required to detect the location of a target which appears on either the left or right side of the screen following the presentation of a central cue (e.g., an arrow; in the case of endogenous cueing tasks). Response times to a validly or invalidly cued target are analysed to determine the reflexive orienting of attention to the central cue. In gaze-cueing, an image of a face or a pair of eyes is presented as the central cue. Participants are required to respond to the target appearing at a location which is either congruently cued (e.g., eyes directed at the target location) or incongruently cued (e.g., eyes directed at an incorrect location). Typically, adults show faster responses to the congruently cued locations, which has been attributed to the natural propensity to orient to gaze direction (see Frischen, Bayliss, & Tipper, 2007 for a review).

Utilising this paradigm in autistic participants, however, has mostly failed to show significant differences in behaviour compared to neurotypical participants (see Birmingham et al., 2012 for a review). For example, in a study by Swettenham, Condie, Campbell, Milne, and Coleman (2003), a photographed face was used as the central cue manipulating the gaze direction to either congruently or incongruently cue the target location. A group of autistic children were compared to a group of neurotypical children matched on age and non-verbal IQ. Results revealed typical gaze orienting in autism with both upright and inverted faces. Interestingly, a few studies compared a gaze-cueing task to an arrow-cueing task in both autism and typical development to investigate if there was evidence for a socially specific difficulty in reflexive orienting of attention in autism. Not only do findings reveal a lack of evidence for attention orienting deficits in autism, they also show lack of evidence for a difficulty in the social domain (e.g., Kuhn et al., 2010; Senju, Tojo, Dairoku, & Hasegawa, 2004).

Gaze-cueing studies have been successful in utilising controlled and objective paradigms to provide us with important insights on the reflexive orienting of attention to gaze shifts. However, these studies lacked the intentional and collaborative component of a social interaction in sending and receiving information (Argyle & Cook, 1976). Hence, these gaze-cueing paradigms could not be used to investigate fully the social nature of joint attention responsivity as a higher domain cognitive ability. This is because they have, potentially, removed most of the elements that make joint attention specifically ‘social’ (Birmingham et al., 2012). This resulted in a compromised ecological validity in favor of experimental control, which could be a major contributor to the inconsistent findings reported so far.

In this chapter, we review some of the most prominent work that has attempted to achieve enhanced ecological validity while maintaining a high level of experimental control. We draw special attention to the importance of utilising interactive paradigms and realistic contexts in better understanding gaze-based joint attention responsivity. We also outline a model of social communication, the Ostensive-Inferential Communication model (Sperber & Wilson, 1986; Wilson & Sperber, 2002), which can be utilised in simulating and investigating ecologically valid and controlled interactive contexts for studying joint attention. This is followed by a review of existing work using one of the most empirically-tested interactive paradigms for objectively measuring joint attention behaviour and corresponding brain processes (Caruana, Brock, & Woolgar, 2015). This paradigm was adapted in the current series of studies to investigate joint attention responsivity in autism. Finally, the rationale for the current work will be presented.

1.1 Gaze-Following Paradigm

In a series of studies, Leekam and colleagues were able to create a structured gaze-following paradigm, which has moved away from reflexive attention orienting gaze-cueing paradigms, to enable investigating joint attention responsivity during a social interaction (Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997; Leekam, López, & Moore, 2000). In the paradigm used by Leekam et al. (2000), an experimenter sat across the table from an infant with two boxes placed on the table between them. One box to the left and the other to the right of the experimenter. Each trial started with the experimenter establishing eye contact with the infant and then shifting her gaze to look at one of the boxes. Once the infant followed the guiding gaze by looking at the cued location, the correct box was opened to show the infant a toy and flickering light. These interactions were video recorded and the time taken for infants to respond were measured from the recordings. This paradigm was beneficial in providing a measure for naturalistic gaze-following behaviour while also enabling the investigation of RJA behaviour as an intentionally elicited, goal directed, attention orienting mechanism. However, it was still limited due to the use of a genuine interaction, which meant it was not possible to maintain the consistency of the presentation of stimuli across trials and participants.

The structured experimental studies investigating joint attention ability in autistic people have most commonly adopted either simple non-interactive paradigms – in which the participant observes and responds to a social stimulus from a third-person perspective –, or natural social interaction paradigms which lack experimental control (Moore & Barresi, 2017; Redcay & Schilbach, 2019; Siposova & Carpenter, 2019). Many researchers have recently emphasised the need to develop more realistic, yet controlled, interactive joint attention paradigms in which the participant evaluates and engages with the social stimulus (or social partner) from a second-person (i.e., we, me and you) perspective (Moore & Barresi, 2017; Schilbach et al., 2013).

1.2 Interactive vs. Non-Interactive: Importance of second-person paradigms

Many researchers in the field of social cognition generally (e.g. Baez, García, & Ibanez, 2016; De Jaegher, 2009; De Jaegher, Di Paolo, & Gallagher, 2010; Gallotti & Frith, 2013; Schilbach, 2010) and joint attention specifically (e.g. Caruana, McArthur, Woolgar, & Brock, 2017b; Gomez, 2005; Mundy, 2018; Redcay & Schilbach, 2019; Siposova & Carpenter, 2019) have highlighted the need for utilising a controlled second-person approach in understanding social cognition and its processes. Recent studies have shown evidence that social cognition operates differently when interacting with others versus simply observing a social interaction or social stimuli (see Redcay & Schilbach, 2019; Schilbach et al., 2013 for reviews).

For instance, Redcay and colleagues (2010) reported increased activation of social brain regions when participants completed a live video feed interaction with the experimenter compared to when they watched a pre-recorded video of the same interaction. In a second experiment within the same study, participants performed a joint attention task by playing a game to catch a mouse hiding in one of four locations. To do this, they needed to follow the gaze of their partner to the target location. This was compared to a task in which participants had to find the mouse on their own by being presented with a cue (i.e., the mouse's tail appearing at the target location). The experimenter was still presented on this task but was not interacting with the participant. Here, greater activation on the joint attention task was found in the right posterior superior temporal sulcus (rpSTS), right temporoparietal junction (rTPJ) and right anterior superior temporal sulcus (raSTS). These findings clearly indicate a difference in brain activation when actively interacting with a social partner compared to passively observing social stimuli or a social interaction between others.

1.3 Controlled Interactive Paradigms of Joint Attention

To this end, several studies have begun to use an interactive approach in the investigation of joint attention (see Caruana et al., 2017b for a review). For example, in another fMRI study, Saito et al. (2010) utilised a real-time video stream to investigate joint attention responsivity in neurotypical people. Participants needed to perform an interactive gaze-cueing joint attention task as well as a non-interactive ball-cueing task. In this study, live video streams of each participant's eyes and eyebrows were presented on a screen for their partner to see during both the interactive and non-interactive tasks. Also, instead of interacting with an experimenter, both interacting partners were participants, in what has been termed a 'dual-brain' approach (Redcay & Schilbach, 2019). Each trial starts by initiating eye contact before either responding to their partner's gaze-cue or a ball-cue in which one of two balls presented at the bottom of the screen changed color. They were either instructed to follow or avoid the target. This study demonstrated a more prominent correlation in brain activity between interacting pairs of participants in the right inferior frontal gyrus as compared to non-paired participants (i.e., randomly selected participants from different interacting dyads). This finding was interpreted as evidence for the specific activation associated with sharing intentions through eye-contact.

These, and similar, studies utilising live-video feed allowed for the investigation of a live interaction while using sophisticated technological tools for more accurate data collection (e.g., Redcay et al., 2013; Redcay, Kleiner, & Saxe, 2012). Although using real-time videos and having two people genuinely interacting with one another does lead to more ecologically valid studies. These studies, however, compromise the level of experimental control by allowing for uncontrollable human factors (e.g., physical appearance, gender) to influence the investigated manipulation. It is also not possible to tightly control the social behaviour (e.g., number of gaze shifts) exhibited by each participant, or the use of subtle non-verbal social cues (e.g., eyebrows). The inability to isolate the manipulation of interest in studies is problematic as it results in measures that are less objective. This, in turn, also leads to challenges in utilising established tools for analysing the data collected and accurately interpreting the results.

Other interactive studies have sought to overcome this limitation by using controlled anthropomorphic virtual agents instead of real-time video. These agents were either controlled by a real person or a gaze-contingent algorithm based on the participant's own gaze behaviour (see Caruana et al., 2017b for a review). In yet another fMRI study, Schilbach et al. (2010) created a socially interactive virtual reality paradigm which provided a promising solution to the ecological validity and experimental control trade-off. This so-called second-person paradigm allowed for a controlled social interaction where participants interacted with an anthropomorphic avatar controlled by a gaze-contingent computer algorithm which they believed to be representing a human partner. Participants' eyes were tracked and the avatar face was presented on the screen along with three squares (to the left, right and above). Participants performed both RJA and IJA trials.

On RJA trials, the avatar looked at one of the squares by performing a single eye-gaze shift to which participants responded by looking at the same square. On matched control trials, participants were required to look at one of the other squares to avoid joint attention. On IJA trials, participants initiated joint attention by looking at one of the squares and the avatar responded by looking at the same location. On control trials, however, the avatar responded by looking at one of the unattended squares. Increased activation was observed in the medial prefrontal cortex (mPFC) on RJA trials compared to IJA trials, over and above activation on the corresponding control trials. Activation in this region has been previously associated with representing the mental states of others (i.e., mentalising; Castelli, Frith, Happé, & Frith, 2002; Castelli, Happé, Frith, & Frith, 2000; Fletcher et al., 1995; Frith & Frith, 2000; Frith, 2001; Gallagher & Frith, 2003; Gallagher et al., 2000) and in self-representation (Cabeza et al., 2004; Heatherton et al., 2006). This finding was interpreted to suggest that RJA is supported by neural mechanisms which enable humans to represent the perspectives of others relative to one's own perspective so that attention can be co-ordinated (Amodio & Frith, 2006; Saxe, 2006; Schilbach et al., 2006).

Given that the paradigms discussed so far have been specifically designed for fMRI studies, they were created in a way that minimises individual behavioural differences to allow for identifying common regions of brain activation. This lead to the use of tasks which were still rather simple and not representative of real-life joint attention. There are two main factors affecting the ecological validity of these tasks: (1) the need for excessive explicit instructions, and (2) the presentation of gaze-cues as single unambiguous cues. For example, in Schilbach et al. (2010), participants were explicitly instructed to begin each trial by establishing eye contact to ensure their partner was ready. They were also informed of their social role (as initiators or responders) and the outcome of joint attention (i.e., joint attention achieved or avoided) using written cues. Additionally, eye-gaze cues were presented as single non-ambiguously communicative gaze shifts that needed to be initiated or followed. This is not representative of real-life joint attention, where social-cues are embedded in a constant stream of ambiguous social information.

For successful responsivity in real-life social interactions, social information needs to be evaluated to identify relevant joint attention bids. Recent evidence suggests that difficulty in social cognition tasks in autism are only notable on tasks requiring spontaneously inferring conclusions using contextual information (Baez et al., 2016; Baez & Ibanez, 2014; Baez et al., 2012). As such, there is a need for experimental joint attention paradigms that capture and test for the influence of contextual factors on joint attention responsivity.

1.4 Gaze in context: Importance of a realistic context

The dual functionality of eye-gaze, in which it is used in both sensing our environment as well as signalling information, makes it a highly ambiguous social cue (Gobel, Kim, & Richardson, 2015; Jarick & Kingstone, 2015; Myllyneva & Hietanen, 2015; Risko, Richardson, & Kingstone, 2016). Recent evidence suggests that understanding and responding to gaze cues is highly context-dependent (Hamilton, 2016). Depending on the context in which it is observed, eye-gaze can be used to signal threat, attentiveness, social dominance or attract social interest (El Zein, Wyart, & Grezes, 2015; Sander, Grandjean, Kaiser, Wehrle, & Scherer, 2007).

The need to spontaneously perceive and integrate relevant social cues during an interaction is critical in supporting our ability to adaptively respond to communicative bids made by others (Klin, 2000). Arguably, humans achieve this by inferring the social meaning of an action using contextual clues (Baez et al., 2012; Ibañez & Manes, 2012). For example, consider a situation where Mary and Ann are working together, and Ann starts complaining to Mary about her boss. While Ann was talking, Mary notices that Ann's boss was approaching from behind her. She flashes her eyebrows and shifts her gaze to glance behind Ann quickly before looking back at Ann's face. In order for Ann to interpret correctly Mary's communicative signals (i.e., intending to inform her that the boss is approaching) Ann had to infer that Mary's eyebrow flash signalled her intention to communicate a relevant information. Critically, this cue alerts Ann to the relevance of Mary's subsequent gaze shift which signals the boss' location in space. Therefore, Mary's gaze shift is only informative if it is evaluated in the context of her other social cues (e.g., eye brow flash) as well as the broader environmental context.

The ability to utilise contextual factors in 'meaning making' during social situations has been suggested to be problematic in autism (Baez & Ibanez, 2014; Vermeulen, 2015). Baez et al. (2012) tested autistic adults on several social cognitive tasks with varying context-processing requirements. The tasks used tested for emotion recognition, theory of mind (ToM), empathy, moral judgment, social norms knowledge, and self-monitoring behaviour in social settings. Eight different tasks and assessments were used. Some of those tasks required participants implicitly to perceive and integrate relevant social elements to solve a social scenario. Others had elements of the social scenario explicitly defined. The results from tasks requiring implicit inference (e.g., empathy for pain task; Couto et al., 2013) revealed a difficulty in inferring the intentionality of actions in autistic people. However, on tasks where explicit and abstract rules were provided (e.g., moral judgement task, where harm and intentionality were explicitly detailed), no significant group differences were reported.

Similar findings have been reported in a number of studies investigating implicit ToM in autism, with evidence for a specific difficulty in spontaneously attributing mental states to others (e.g., Klin, 2000; Senju, Southgate, White, & Frith, 2009). This difficulty is evident even when participants' performance on explicit ToM tasks, which involved explicit verbal instructions and scenarios, was not significantly different to age matched neurotypical controls (see Sodian et al., 2015 for a review).

In a subsequent paper, Baez et al. (2016) noted that most experimental tasks currently utilised in investigating social cognition fail to include the essential influence of implicit contextual information processing. In their paper, they point out the need for more realistic, ecologically-valid paradigms that "control for context-dependent levels in social cognition tasks" (Baez et al., 2016, p.392) including 'context-free' and 'context-rich' elements to investigate properly the effect of context in clinical populations such as autism. In real-life joint attention episodes, utilising contextual clues to make inferences about the communicative intentions of a social partner as well as the social meaning of their cues is fundamental for a successful interaction. Hence, investigating this ability without accounting for the effect of contextual information processing could result in unrealistic experimental studies with compromised ecological validity.

To identify the specific contextual factors which influence responsivity during gaze-based joint attention, we can draw upon empirically-supported theoretical accounts of social communication. The next section outlines an important model of social communication which was used as a framework for deriving specific hypotheses to test in the current series of studies.

1.5 Relevance Theory: Ostensive-inferential communication model

In an influential theory of social communication, the Relevance Theory, Sperber and Wilson (1986) postulated that human cognition is driven by the natural tendency to search for relevance when processing any form of input (e.g., a thought, a sight, a sound, an utterance).

According to this theory, an input is considered relevant when (1) it can be associated with prior information someone has and (2) it leads to a conclusion that is meaningful and useful to them. The authors also proposed a specific model for relevance processing during social communication, which has been highly influential in the field of linguistic pragmatics (Scott-Phillips, 2018). This model is based upon an idea previously proposed by Grice (1957), who suggested that for a successful social communication to take place, a communicator is expected to express two intentions: an informative intention (i.e., the intention to inform the receiver of relevant information), and a communicative intention (i.e., the intention to explicitly inform the receiver of their intention to communicate).

Sperber and Wilson suggest that sharing *communicative intentions* with the intended receiver assists in raising their expectation of receiving a relevant information (Wilson & Sperber, 2002). This is thought to result in capturing and orienting the receiver's attention to relevant contextual clues which allows the receiver to infer the communicator's meaning in the most efficient way possible. 'Ostensive signals' (e.g., Mary's eyebrow flash in the example above) are used to convey communicative intent to the intended receiver. When such signals are successful in supporting social inferences by the receiver in an interaction, they are said to have 'optimal relevance' – and inform precise and predictable expectations of relevance, unlike any other non-communicative behaviours (Wilson & Sperber, 2002). Furthermore, a social cue is of optimal relevance when it is the most relevant social cue in signalling the communicator's intent and is detected by the receiver with the least processing effort possible.

1.6 Studies investigating the Ostensive-Inferential Communication Model

The ability of children to use ostensive signals in inferring the communicative and informative intentions of a social partner has been previously investigated, with evidence in support of this model. For example, in a study by Behne, Carpenter, and Tomasello (2005), a hiding-finding game was used to investigate the ability of children aged 14, 18 and 24 months to infer the communicative intent of an adult from ostensive signals. A toy was hidden in one of two containers by an experimenter and the correct location was indicated using either (1) pointing paired with ostensive gazing, or (2) ostensive gazing only as communicative cues. During ostensive gazing, the experimenter alternated her gaze between the cued container and the child while raising her eyebrows to express communicative intent. The children were then prompted to find the toy by grabbing the correct container. Performance on these conditions was compared to two control conditions. Here, instead of using ostensive signals, the experimenter looked at the hiding location 'absent-mindedly' (i.e., unfocused gaze with neutral facial expressions) and had her hand held in the same position as in the pointing condition but did not include communicative cues (i.e., pointing or ostensive gazing). Children as young as 14 months old were able to infer the communicative intent from ostensive signals in conditions (1) and (2). They were also able to understand the experimenter's intention to help them find the toy by inferring the relevance of their ostensive behaviour. This allowed them to correctly determine the target location only on the communicative conditions while performance on the control condition was not greater than chance.

A few studies have attempted to investigate the ability to intentionally elicit gaze following in infants by manipulating the presence of ostensive signals, such as eye contact or infant-directed greeting (Farroni, Mansfield, Lai, & Johnson, 2003; Senju & Csibra, 2008). Evidence reveal that intentionally directing the attention of infants' gaze was only possible in the presence of ostensive signals. Senju, Csibra, and Johnson (2008) investigated the effect of eye contact as an ostensive signal on associating a relationship between gaze direction and object location in 9-month-old infants. Their results suggest that eye contact was necessary to deduce a relationship between gaze shifts and objects. Other studies have also shown that, in ostensive contexts (e.g., in the presence of eye contact, pointing, eyebrow flashes), infants do not only follow gaze but also expect to find a relevant object of reference (Behne et al., 2005; Csibra & Volein, 2008; Gliga & Csibra, 2009). Several studies have suggested that perceived eye contact has been found to modulate the processing of accompanying sensory information in a phenomena that has been termed the ‘eye contact effect’ (Senju & Johnson, 2009b). These results emphasise the importance of eye contact as an ostensive signal in intentionally conveying the intention to communicate and its influence on orienting the attention of infants to a relevant object.

Böckler, Knoblich, and Sebanz (2011) utilised a third-person paradigm to investigate the effect of observing eye contact between two people on the gaze following behaviour of a third person (the observer). In this study, participants observed two faces either look at each other or away from one another before shifting their gaze together either to look at the target location or toward an incorrect location. A target was then presented at one of the two locations and participants were required to respond by pressing one of two keys to denote the location. Participants were faster to respond to correctly-cued trials only when the faces shared mutual gaze. In a subsequent study, the same experiment was conducted to investigate this effect in autistic people (Böckler, Timmermans, Sebanz, Vogeley, & Schilbach, 2014). Results show that, unlike typical adults, autistic participants showed no effect of eye contact on their gaze following behaviour. Similar findings in autism have been reported with evidence for a lack of effect of perceived eye contact on responsivity (see Senju & Johnson, 2009a for a review). These studies, however, have only investigated the effect of ostensive signals using a third-person as opposed to a second-person approach. It is, hence, still unclear if eye contact is evaluated as an ostensive signal by autistic people differently during a social interaction, and how this affects joint attention responsivity.

Based on the ostensive-inferential model (Sperber & Wilson, 1986; Wilson & Sperber, 2002), our natural tendency to associate optimal relevance to ostensive signals allows us to detect social communication bids that are relevant to us (e.g., joint attention bids). This, in turn, assists us in distinguishing between the relevant and irrelevant social information we are constantly surrounded by in social settings. This ability is crucial for successfully achieving joint attention in real-life social situations. To properly understand gaze-based joint attention abilities, we would benefit from utilising the ostensive-inferential communication model as a framework for deriving hypotheses and incorporating context in designing controlled second-person interactive paradigms.

The first steps towards achieving this goal have been taken by Caruana et al. (2015), through creating a context-dependent interactive virtual reality joint attention task. In the next section, we discuss the findings and limitations of the previous work using this paradigm followed by the rationale for this study.

1.7 Review of previous work

A new interactive joint attention task, the 'Catch the Burglar' game, was developed by Caruana et al. (2015) to incorporate a realistic context in second-person joint attention paradigms. In this paradigm, participants played a cooperative game in which they interacted with a computer-controlled avatar. They were made to believe that the avatar was controlled by another experimenter in an adjacent laboratory. In this task, participants were presented with six houses (three on top, three on bottom; see Figure 1). They were informed that they needed to work collaboratively with their partner to search for a burglar that was hiding in one of these six houses, and whoever finds the burglar first will need to guide the other person to the correct location. They were assigned to search three houses (on top) and their partners the three others (on bottom). Participants were provided with no explicit instructions as to how they should collaborate with their partner to complete the task. Hence, they were required to evaluate the cues in a rather realistic manner.

On IJA trials, participants found the burglar in one of their houses and were required to capture the attention of their partner and guide them to its location. To do this successfully, they needed to establish eye contact with the avatar before shifting their gaze to the correct location. On RJA trials, participants did not find the burglar in their houses and needed to wait for their partner to complete their search before being guided by them. Here, the avatar established eye contact with the participant before guiding them to the location of the burglar. The addition of the initial search phase provided a natural and realistic context where participants needed to evaluate eye-gaze behaviour and identify the relevant guiding cues from a stream of irrelevant, non-communicative, eye movements. That is, this required

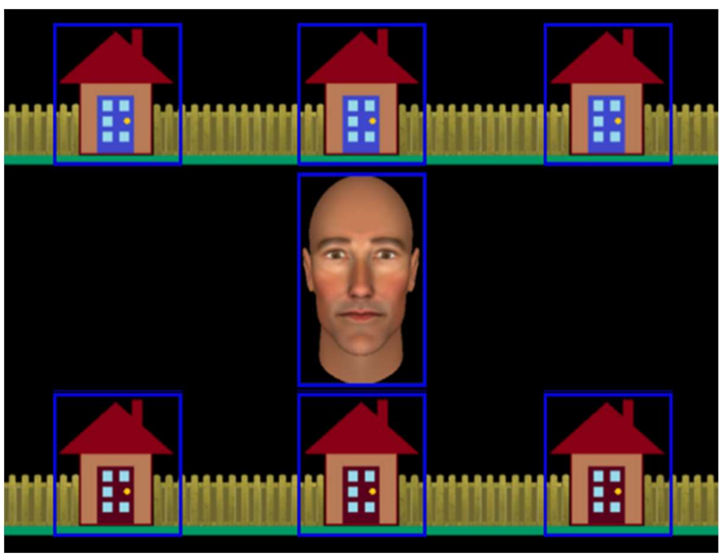

Figure 1. Stimuli used in the interactive joint attention task, including the central avatar and the six houses in which the burglar could be hiding. Gaze-related areas of interest (AOIs), are represented by blue rectangles. These were not visible to participants.

participants to evaluate the avatar's eye contact as an ostensive signal to differentiate the preceding non-communicative gaze shifts from the subsequent 'communicative' gaze shift.

A control condition – a matched non-social task – was also utilised, in which participants completed the same task except they were told that they were completing a computer simulation instead. This condition was employed to rule out the potential effects of non-social task demands, such as attentional orienting and oculomotor control. In this condition, the avatar was presented on the screen with closed eyes with a grey fixation point between its eyes. On IJA trials, participants were explicitly instructed to look at the grey fixation point upon finding the burglar, wait for it to turn green (analogous to eye contact) and then look back at the location of the burglar to catch it. On RJA trials, after completing their search and not finding the burglar, participants fixated the grey fixation point before it turned green (analogous to eye contact). Then a green arrow extended from the fixation point (analogous to communicative gaze) and guided them to the correct location.

This paradigm has been previously used to investigate the neural correlates common to both RJA and IJA behaviours with evidence for a right-lateralised frontotemporoparietal network supporting both functions (Caruana et al., 2015). It has also been used to investigate joint attention behaviour in autistic adults (Caruana et al., 2018), who made more errors than typical adults when responding to Eyes in the social condition, but not when responding to Arrows in the control condition. Furthermore, participants completed two blocks of the task, in which social and non-social trials were interleaved. Whilst autistic individuals were slower to respond to gaze-cued joint attention bids compared to typical adults on the first block, responsivity increased and performance was commensurate to typical adults on the second block. Again the same pattern was not observed for non-social arrow trials in which participants were faster to respond to Arrows overall, and performance was commensurate between groups from the beginning of the task. Together, these findings suggested that the observed differences in the autism group resembled a challenge specific to the social domain.

The socially-specific challenge in autism has been attributed to a difficulty in differentiating between the relevant communicative gaze-cues and the irrelevant, non-communicative gaze shifts during the search-phase. This interpretation was supported with evidence from the subjective interviews indicating that autistic participants found it hard to understand the meaning of the eye movements. However, since this study did not manipulate the different components of the contextual information included in the search-phase, the specific factors that contribute to this challenge in autism were not identified.

In a subsequent study, Caruana, McArthur, Woolgar, and Brock (2017a) investigated the specific effect of adding the ‘search-phase’ on the responsivity of typical adults. In this study, participants performed a task with a search-phase (context-dependent) and another without a search-phase (context-free). Results suggested that participants were significantly slower on the ‘Search’ than the ‘NoSearch’ task only with the Eyes stimulus condition. They were also found to be slower when responding to Eyes than Arrows on both the Search and NoSearch contexts. These results were interpreted as evidence for the influence of the intention monitoring processes on joint attention responsivity. However, one limitation of the paradigm used in these studies was the use of a non-social Arrows control condition which did not fully match the avatar’s eye-gaze shifts during the search-phase on the social Eyes condition. On the non-social condition, when participants performed the search, no stimuli were updated until the search was complete. On the social condition, however, the avatar’s gaze shifts were updated continuously for their partner to appear to be searching through his houses. Therefore the reported effects in this study, and the previous study in autism (Caruana et al., 2018), could have simply been due to the non-social Arrows condition being less demanding since it did not require evaluating spatial information in context to determine their relevance.

This limitation was addressed in a recent study where a better-matched control was implemented, and the arrow stimuli were programmed to mimic the avatar's eye movement behaviour during the search phase (Caruana et al., in prep.). By doing this, the previously observed effect of stimulus on the Search condition as well as the context effect on the social (Eyes) condition were no longer observed in typical adults. We also observed that participants made more errors on the Search than the NoSearch context conditions with both Eyes and Arrows stimulus conditions. Their reaction times to joint attention guiding bids, however, were found to be slower on the Search condition only with the Arrows condition, but not Eyes. Hence, the cost of evaluating contextual clues was found to be larger for Arrows than Eyes, which suggests that typical adults have a relative advantage for evaluating gaze-based social contextual information and identifying relevant communicative cues.

These findings emphasised the importance of implementing a well-matched control condition that accounts for the effect of contextual processing of spatial information. This matched paradigm allowed us to investigate whether the demonstrated effects were specific to the social domain, or were rather domain-general effects associated with the ability to determine spatial cue patterns. The reported findings suggested a relative advantage for Eyes which could be due to eye contact serving as an ostensive signal.

In a second experiment of the same study (Caruana et al., in prep.), we manipulated the informative nature of the contextual clues presented during the search-phase to investigate its effect on adults' RJA behaviour. Here we used the same search context condition of the 'Catch the Burglar' game. On this condition, the avatar was programmed to randomly determine the final house looked at during the search-phase before establishing eye contact and guiding the participant on RJA trials (Random Search). This condition was compared to another condition where the avatar was programmed to look at the burglar's location last before guiding the participant to that same location. Therefore, on these trials, the avatar's non-communicative searching gaze was predictive of the target location (Predictive Search). Participants made significantly more errors on the Random Search than the Predictive Search context and were significantly faster to respond on the Predictive Search than the Random Search with both stimuli (Eyes and Arrows). Participants were also significantly faster to respond to the Predictive Arrows than the Predictive Eyes. Importantly, results also revealed a significant context by stimulus interaction, suggesting a relative advantage for evaluating the random contextual information in the social Eyes condition when compared to the non-social Arrows condition. The relative advantage for Eyes reported by both studies has been attributed to the ability to infer the communicative intent of a social partner from the direct-gaze stimulus (i.e., the ostensive signal), in line with the ostensive communication model (Sperber & Wilson, 1986).

1.8 The current study

There were two key limitations to the paradigm used in the latest studies, which limit the sorts of conclusions that can be drawn. The first limitation was that the green fixation point stimulus in the non-social Arrows condition, which was meant to be analogous to direct gaze, did not match the effect of eye contact in capturing attention (Caruana et al., in prep.). On Arrow trials, the green fixation point was followed by a green arrow that extended from the same green point. This resulted in a relatively subtle transition between 'searching' and 'guiding' arrow shifts which were intervened by the fixation point. This contrasts with the arguably more obvious transition between searching and guiding gaze, which was intervened by direct gaze. As such, it is possible that the previously reported advantage for eye gaze was not related to the ostensive nature of eye contact, but rather the more obvious visual contrast between direct and averted gaze. To remedy this, in the current work, a yellow fixation point stimulus replaced the green fixation point on Arrow trials (see Figure 2) to better match for the effect of eye contact in capturing attention.

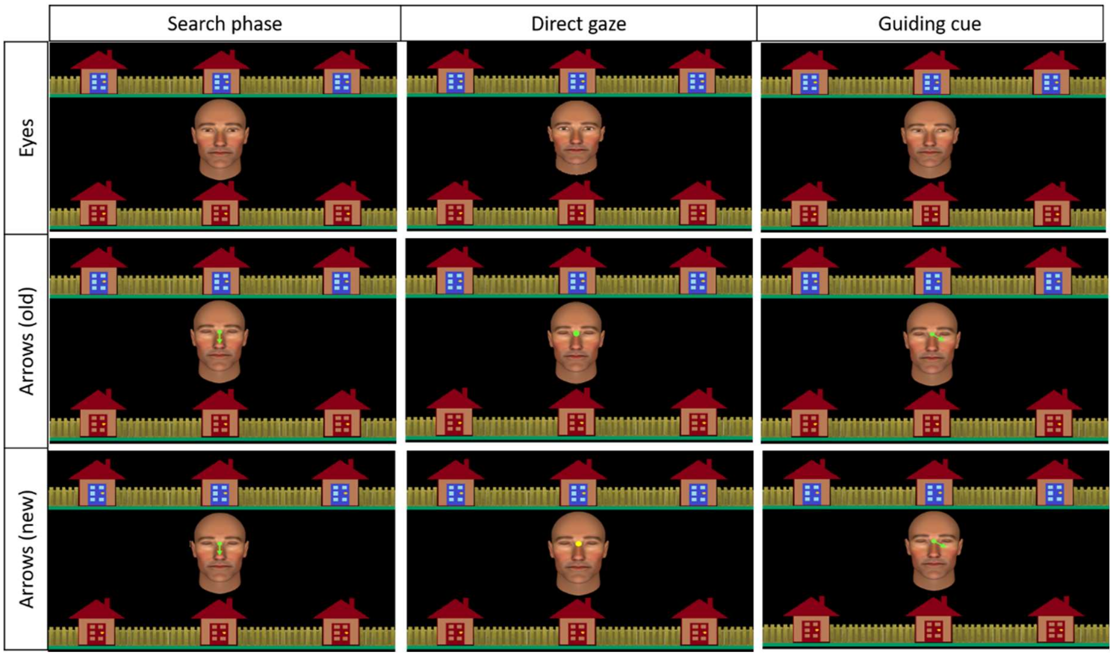

Figure 2. Illustration of the yellow fixation point used in this study.

The second limitation was that two separate experiments were conducted to compare (1) Random and NoSearch, and (2) Random and Predictive Search contexts with two distinct samples. This between-participants design does not allow for the comparison of both context-dependent conditions (i.e., Random and Predictive Search) to the context-free condition (i.e., NoSearch) as the baseline. Therefore, the overall effect of adding a predictive context on the ability to predict the direction of the guiding cue as compared to the context-free condition (NoSearch) was not clear. To understand this issue further, in Chapter 2 we conducted an experiment in which adults' performance was compared across all three context conditions using a fully within-participants design (i.e., Random Search, Predictive Search, NoSearch). In Chapter 3, we take the first step towards investigating the effect of contextual information on the ability of autistic adolescents and young adults to respond to gaze-based joint attention. To do this we compared responsivity on the Random Search to NoSearch context conditions only. In both experiments, social Eyes and non-social Arrows conditions were compared to determine whether the effects were specific to the social domain.

Chapter 2: Effect of context in adults (Experiment 1)

2.1 Introduction

In this study, we aimed to understand the ability to identify communicative (i.e., relevant) gaze cues when embedded in a realistic context of non-communicative (i.e., irrelevant) gaze behaviour and the effect of this ability on joint attention responsivity. We did this by utilising the 'Catch the Burglar' game from Caruana et al. (in prep.) to investigate the effect of three non-communicative gaze contexts (i.e., Random Search, NoSearch and Predictive Search) using a fully within-participants design. We analysed participants' ability to respond accurately to their partner's communicative cues by looking in the correct direction to successfully achieve joint attention. We also analysed their SRTs on trials where they accurately responded to communicative gaze cues.

2.1.1 Hypotheses

Based on previous findings from Caruana et al. (in prep.), and consistent with the ostensive-inferential communication model (Sperber & Wilson, 1986; Wilson & Sperber, 2002), we hypothesised that ostensive signals should show a unique advantage by assisting participants to decipher between relevant communicative gaze-cues and irrelevant gaze-shifts. Therefore, we expected to find the following:

Accuracy. Participants would make more errors on the Random Search context than both the NoSearch and Predictive Search contexts.

SRT. Participants would be slower to respond on the Random Arrows condition but not the Random Eyes when compared to the NoSearch condition. We also expected that participants would have faster reaction times on the Predictive Search context when compared to both NoSearch and Random Search conditions. Finally, we expected to find a relative advantage for identifying relevant gaze cues than relevant arrow cues when comparing both the Random Search and Predictive Search contexts to NoSearch on SRT.

2.2 Methods

2.2.1 Ethics Statement

All procedures implemented in this study were approved by the Macquarie University Human Research Ethics Committee (ID: **3775**). All participants gave written, informed consent to take part in this study prior to participation.

2.2.2 Participants

Thirty-one adult participants were recruited from a pool of undergraduate psychology students at Macquarie University and were given course credit for their time. All participants reported normal or corrected-to-normal vision with no history of neurological injury or impairment. Five participants were excluded due to technical failure of the eye-tracking calibration (n=3) or did not believe the deceptive cover story (n=2). Therefore, the final sample included 26 participants (M_{age}= 18.96 years; *SD*= 1.34; 18 females).

2.2.3 Stimulus and Apparatus

Participants were seated at a table with a chin and forehead rest installed to stabilise their head movements and standardise the screen viewing distance. Participants played a cooperative game with an on-screen avatar. The experimental stimuli were presented using Experiment Builder 1.10.165 on a 27-inch AOC monitor (display size: 59.8 cm x 33.6 cm; resolution: 1920 x 1080 pixels; refresh rate: 144 Hz) positioned 80 cm away from the participant. A remote desktop-mounted Eyelink 1000 (SR Research Ltd., Ontario, Canada) was used to record eye-movements from the right eye at a sampling rate of 500Hz. Before starting each block, a 9-point eye-tracking calibration and validation was implemented.

During the task trials, an avatar face, presented as an anthropomorphic animated face subtending 6.08° x 3.65°, was displayed in the center of the screen surrounded by six houses, each subtending 3.58°, as shown in Figure 1. There were seven invisible gaze-related areas of interest (AOIs; see Figure 1) defined around the avatar's face and each of the six houses which were used by our gaze-contingent algorithm during the experiment and for subsequent data processing (see Caruana et al., 2015, for detailed description).

2.2.4 Design and procedure

Participants were told that they would be playing a collaborative game with two different members of the research team, named Alan and Tony, who would be interacting with them from an adjacent laboratory. They were told that their partner's eyes would be recorded using an eye-tracker and used to control the eye movements of the avatar they saw on their screen. They were also told that they would control an avatar on Alan/Tony's screen in the same way. In reality, the avatar's eye movements were controlled by a gaze-contingent algorithm.

Each trial starts with the participant searching for the burglar in the houses with blue doors located at the top of the screen, while the avatar searched the red doored houses at the bottom of the screen. We refer to this as the 'search-phase' of trials as illustrated in Figure 3. To search their allocated houses, participants had to look at a house before the door would open and either reveal the burglar or an empty house. To help introduce variability in the spatial sequence of participant's search behaviour, some trials started with one or two houses already open and empty. The sequence and number of open houses were systematically varied across trials to help justify Alan's unpredictable search behaviour by making it more realistic, and resulting in a more believable cover story.

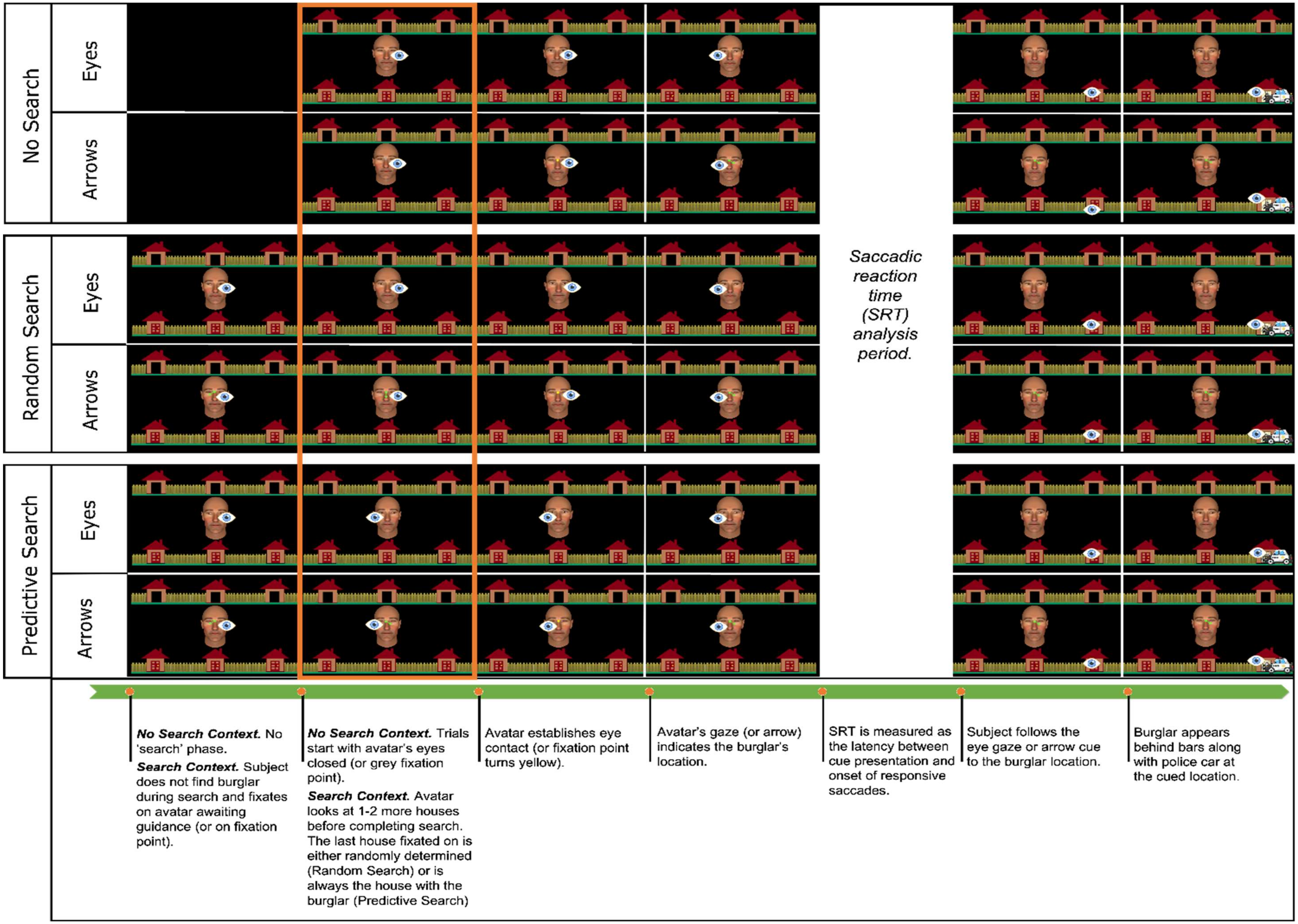

Figure 3. Schematic representation of trial sequence by condition from the completion of the Search phase.

On responding trials, participants did not find the burglar in any of their houses and had to wait for their partner to complete his search and guide them to the correct location. On initiating trials, participants found the burglar in one of their houses and had to capture their partner's attention by establishing eye contact before guiding to the location of the burglar. They were told that for them to successfully catch the burglar they need to be both looking at the target location. Participants were not given explicit instructions as to how they could guide or be guided by their partner. Once joint attention was achieved, the burglar appeared behind bars with a police car at the correct location to provide positive feedback.

In this study, we manipulated both the context preceding the joint attention bid as well as the stimulus used, yielding five conditions: three *context* conditions (No Search, Random Search and Predictive Search) and two *stimulus* conditions (Eyes and Arrows). Participants had to complete six blocks of trials of each context by stimulus combination, with 30 responding trials and 30 initiating trials per block. Although the main interest of this study was on responding trials, keeping the initiating trials was important to maintain the context realism by maintaining the reciprocal nature of a social interaction (Argyle & Cook, 1976). It was important for the task to be an active social collaboration in which the cues were evaluated for their relevance in a realistic manner. Participants needed intuitively to determine their role in the interaction (i.e., initiating or responding) and also needed to capture the attention of their partner to communicate with them and establish joint attention in a reciprocal manner. This context would not be possible if we only had responding trials, since the participant will be attending to the avatar while passively waiting for their partner to complete their search and guide them to the target location.

The stimulus conditions of each context were always administered consecutively to minimise the switching between each search condition (i.e., Eyes Random Search, Arrows Random Search, Eyes Predictive Search, Arrows Predictive Search, Eyes No Search, Arrows

No Search). Within each block, trial order was randomised to ensure that the location of the burglar and the number of gaze shifts made by the avatar were not conflated with order effects.

Participants could make four types of errors on each trial. A Search error refers to trials where participants spent more than 3000 ms looking away from the avatar or their houses during the search-phase. On these trials, participants got a message that says “Failed Search” to prompt them to search through their houses at the beginning of each trial. A Timeout error refers to trials in which participants took more than 3000 ms to respond to the relevant gaze-cue. Location errors refer to trials in which participants respond by looking at an incorrect location. On both Timeout and Location errors, the burglar appeared in red at the target location to provide negative feedback. Finally, Calibration error refers to trials which were interrupted to prompt the recalibration of the eye-trackers. Error trials were excluded from subsequent analyses of SRT data.

Context Conditions

Random Search condition. In this condition each trial started with participants having to search for the burglar. During this search phase, Alan’s gaze would shift each time the participant searched a house, appearing to search his houses in a randomised order. Participants were not able to see the contents of Alan’s houses while they were both searching. On responding trials, where participants did not find the burglar in any of their houses, they needed to fixate back on the avatar’s face and wait for Alan to complete his search and guide them to the target location. Once they fixated back on the avatar’s face, Alan searched 1-2 more houses where the last house he looked at before establishing eye contact was randomly determined and was not predictive of the burglar’s location. Alan then looked at the participant to establish eye contact and initiated joint attention by averting his gaze to the correct house. To catch the burglar, participants needed to respond by looking at the cued house.

Predictive Search condition. This condition was similar to the Random Search condition in that trials also started with the search-phase component. However, participants were told that they would be playing the game with another member of our research team named 'Tony'. Tony was presented on the screen by the same avatar face used for 'Alan'. The difference between the Predictive and the Random conditions was that Tony's final gaze shift during the search-phase was always directed to the target location. This made his final gaze shift before establishing eye contact predictive of the location of the burglar. Given that the avatar's behaviour is systematically different across the Random and Predictive contexts, it is possible that some participants might notice the systematic difference. Indeed 10 of the 34 (29.4 %) participants noticed this in our previous study (Caruana et al., in prep.). As such, we told participants that they were interacting with different people across these two contexts implicitly to provide a realistic explanation for this systematic difference. Importantly, participants were not explicitly informed of this manipulation.

No Search condition. In this condition, participants were told that they were interacting with Alan again but there was no search-phase at the beginning of each trial. Instead, the trials started with the avatar's eyes being closed and the participant's allocated houses being open. On initiating trials, the participants could see the burglar in one of their houses at the top of the screen and had to look at the avatar's face until it opens its eyes, to establish eye contact, before guiding Alan to the correct location. On responding trials, after establishing eye contact, the avatar shifted its gaze towards the burglar's location and participants needed to follow the avatar's gaze to catch the burglar successfully.

Stimulus Conditions

A matched control condition was completed as separate blocks for each context condition. This was implemented to control for non-social task demands, such as attentional, oculomotor and inhibitory control. This resulted in two stimulus conditions (Eyes, Arrows) utilising the same gaze-contingent algorithm for presenting stimuli in both conditions. For the arrow condition, participants were informed that they were completing a computer-simulated version of the task in which a computer-controlled arrow stimulus was used to guide them to the correct location. The avatar's face with closed eyes remained on the screen throughout this condition to match the visual context between both stimulus conditions. At the beginning of each block on Search conditions, and each trial on NoSearch condition, a grey fixation point subtending a visual angle of 0.29° was presented in between the avatar's eyes and was analogous to the closed eyes avatar of the eye gaze trials. This fixation point then turned yellow, to match the avatar's direct eye gaze. This was followed by a green arrow extending from a green central point subtending a visual angle of 1.08°. This was analogous of the avatar's averted gaze. For initiating trials, participants were instructed to look at the yellow fixation point and then look back at the correct location to catch the burglar. Similarly, on responding trials, they were informed that they needed to look at the fixation point and an arrow stimulus will guide them to the location of the burglar. During the search phase of the Random and Predictive Search conditions, the arrow stimulus was updated to point at different houses to match the avatar's searching behaviour on the social condition. Therefore, participants had to evaluate the arrows to determine the relevant guiding cue in a similar way to that needed during the social condition with Eyes stimulus. Once participants completed their search and looked back at the central area of interest (AOI), the arrow stimulus pointed at 1-2 more houses before being replaced by the yellow fixation point, analogous of eye contact. This was then followed by a single green arrow pointing towards the target house which participants needed to follow to successfully catch the burglar. This component was missing in the previous paradigms and was only implemented in the most recent study (Caruana et al., in prep.) with the exception of the yellow fixation point which previously remained green.

2.2.5 Statistical Analyses

Interest area and trial reports were exported using DataViewer software (SR Research Ltd., Ontario, Canada) to analyse the accuracy and SRT data. For accuracy, Calibration and Search errors were removed before analysing the remaining trials for the proportion of correct trials. This was done because these errors occurred before the relevant gaze or arrow cue was presented and, hence, do not represent inaccurate responding. For SRT analyses we only included correct trials. Trials in which participants responded faster than 150 ms were also excluded as these were likely to be anticipatory responses (Carpenter, 1988). Raw eye-tracking data was screened and analysed using R using a custom script. The full data set and R code with the analysis outputs and annotated code descriptions are available on the Open Science Framework (https://osf.io/7hdj8/).

Logistic and linear mixed random effects (LME) were used to analyse accuracy and SRT respectively. Specifically, we wanted to evaluate evidence for effects of Context and Stimulus and their interaction. The maximum likelihood estimation method was implemented in these analyses using the *lme4* R package (Bates & Sarkar, 2005) and p-values were estimated using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2015). We used LME modeling over traditional ANOVA analyses because this allows for the estimation of models which account for both subject and item-level random effects when estimating fixed effect parameters. Furthermore, unlike ANOVAs for aggregated data, LME models are also robust to missing data and are suitable for datasets with unbalanced observations in each condition, since each trial – rather than each participant – is treated as a unique observation (Quené & Van den Bergh, 2004, 2008). This is particularly important when comparing reaction time data between participant groups that are likely to differ in their accuracy rates – and therefore differ in the amount of trials that are fed into the reaction time analyses per subject and group. However, we have also conducted traditional ANOVA analyses – which are included in accompanying R code and output (https://osf.io/7hdj8/) – to facilitate comparison with earlier work and across both analysis approaches.

We were interested in investigating the main effect of stimulus (Eyes, Arrows), context (Random, NoSearch, Predictive), and their interaction. To do this we used the 'successive differences contrast coding' method within the 'MASS' package in R (Ripley et al., 2013). This contrast method estimates effect parameters by sequentially comparing each level of context with the next level specified in the model. This method was used to provide parameter estimates for the overall effect of stimulus as well as the context effects between (1) NoSearch and Random Search, and (2) NoSearch and Predictive Search. Of critical interest in the current study, we were interested in testing the interaction of context and stimulus effects. As such, parameter estimates were also obtained for the context-by-stimulus interaction in (1) and (2) above. We were also interested in investigating the effect of context between the Predictive and Random Search conditions as well as their interaction with stimulus. However, this contrast could not be estimated using the predefined successive differences contrast coding method. Therefore, we used the 'emmeans' package to manually define these missing comparisons (Lenth, Singmann, Love, Buerkner, & Herve, 2019). Finally, we ran post-hoc analyses for accuracy and SRT data to assist in interpreting significant stimulus-by-context interaction effects. An FDR correction was then applied to these post-hoc contrasts to confirm significance after correcting for multiple comparisons (Benjamini & Hochberg, 1995). Any discrepancies have been reported in the results section below and detailed analyses with corrected p-values are included in the accompanying R code and output (https://osf.io/7hdj8/).

Accuracy and SRT models were defined with maximally-defined random-factor structures, including random intercepts for trial and by-subject random slopes for the intercept and fixed effects (Barr, Levy, Scheepers, & Tily, 2013). For SRT analyses, the residuals of the raw data violated the normality assumption and hence data were transformed using an inverse transformation. The normality assumption was confirmed after applying the transformation (details can be found in accompanying R code and output (https://osf.io/7hdj8/); see Balota, Aschenbrenner, & Yap, 2013). All analyses had a significance criterion of $\alpha = 0.05$.

For estimating effect-size, the chi-squared goodness-of-fit tests were performed comparing a number of mixed random-effects models using Chi-square likelihood ratios to quantify the contribution of each fixed effect as well as the interaction parameter to the model fit (Johnston, Berry, & Mielke Jr, 2006). Unlike traditional measures of effect-size (e.g., r^2), this method provides an estimation of the variance explained by each fixed effect whilst also accounting for variance independently explained by the specified random effects. This approach has been established in a previous study using a similar paradigm to investigate joint attention responsivity in schizophrenia (see Caruana, Seymour, Brock, & Langdon, 2019). For each analysis, a model containing only the maximally-defined random effects structure was defined (i.e., without including fixed effect factors). Then a series of models were defined, adding one of our fixed-effect parameters at a time (see accompanying R code for a detailed description). We compared between these models using the ‘anova’ function in R to produce a Chi-square likelihood ratio. These ratios indicated the extent to which each parameter improved the model’s fit.

2.3 Results

Accuracy. First, we investigated whether and how the presence of random and predictive spatial signals differentially affected participants' ability to respond correctly to subsequent eye gaze and arrow cues. The accuracy data for context and stimulus conditions are illustrated in Figure 4. On average, participants made significantly more errors on the Random than the NoSearch context conditions. There was neither a significant main effect of stimulus nor a significant effect of context on accuracy between the Predictive and NoSearch contexts. There was also no stimulus-by-context interactions between NoSearch and either the Random or Predictive conditions. Using emmeans to compare the results on the Random and Predictive Search contexts revealed a significant context effect as well as a stimulus-by-context interaction. Post-hoc analysis indicated that this interaction was characterised by a larger effect of context for Arrows than Eyes. In total, only 7% of trials were error trials, with the majority being Location errors (M = 3.17% of trials, SD = 11.74). This was followed closely by Search errors (M = 2.02%, SD = 10.78) and Timeout errors (M = 1.67%, SD = 6.34). Descriptive statistics of estimated fixed effect parameters are summarised in Table 1.

Table 1. Estimated fixed effect parameters for Accuracy (Experiment 1).

Fixed effect	β-coefficient	Standard Error (SE)	z-value	p-value
Context				
NoSearch-Random	2.143	0.424	5.050	< .001***
Predictive-NoSearch	-0.759	0.587	-1.293	0.196
Predictive-Random	-2.768	0.697	-3.971	< .0001***
Stimulus				
Arrows-Eyes	-0.070	0.366	-0.191	0.848
Stimulus*Context				
NoSearch-Random	-0.643	0.501	-1.281	0.200
Predictive-NoSearch	-0.172	0.592	-0.290	0.771

Random-Predictive	-0.815	0.407	-2.001	0.045*
Follow-up comparisons				
Random-Predictive (Arrows)	-1.791	0.404	-4.436	< .0001***[a]
Random-Predictive (Eyes)	-0.977	0.403	-2.421	0.015*[a]

Note. *$p < .05$, ** $p < .01$, *** $p < .001$, [a] Uncorrected *p*-values. However effects remain significant after applying an FDR correction for multiple comparisons (all *p*s < 0.02).

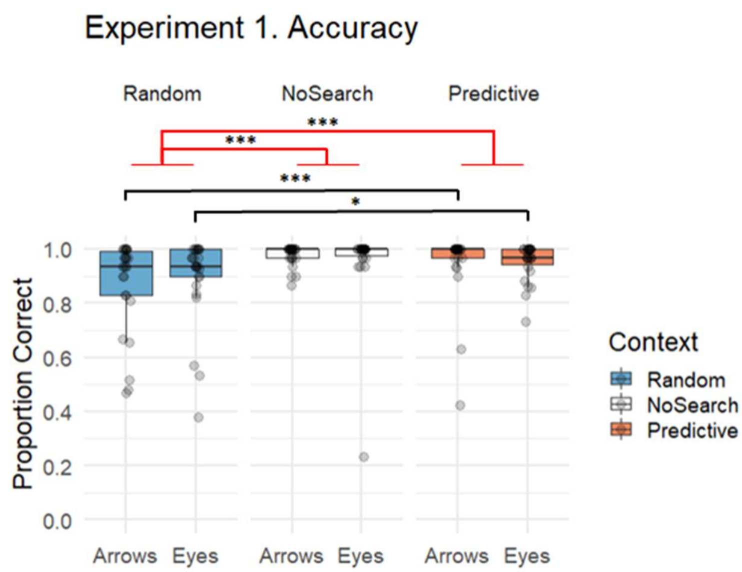

Figure 4. Boxplot with individual data points illustrating the proportion of correct trials by context (Random Search, No Search) and stimulus (Arrows, Eyes). In all boxplot figures, whiskers extend (as in a conventional Tukey's boxplot) 1.5 times the length of the box (i.e., the interquartile range of the 1st and 3rd quartiles). Significant effects of context are illustrated in red while post-hoc contrasts are shown in black (* $p < .05$, *** $p < .001$).

SRT. Next, we investigated whether and how the presence of random and predictive spatial signals differentially influenced the speed with which participants were able to prepare saccadic responses to eye gaze and arrow cues. The SRT data for context and stimulus conditions are illustrated by Figure 5. Results show a significant main effect of stimulus where participants were significantly faster in responding with Arrows stimuli than Eyes.

They were also significantly faster when responding during Predictive trials than NoSearch trials. However, there was no significant effect of context between Random and NoSearch contexts. There were also significant stimulus-by-context interactions when comparing each of the Random and Predictive contexts to the NoSearch context. The context effect and the stimulus-by-context interaction between Random and Predictive contexts were produced using emmeans and revealed significantly slower responses on the Random context than the Predictive context but not a significant interaction.

We also used emmeans to conduct post-hoc pairwise comparisons of interest using custom contrasts to help understand the significant interaction effects. This analysis revealed a significant effect of context between Random and NoSearch with Arrows but not Eyes. There was, also, a significant effect of context between Predictive and NoSearch with a larger effect in Eyes than Arrows. Additionally, the effect of stimulus was only significant in the NoSearch context but not in Random or Predictive. Mean SRT are summarised by condition in Table 2 and descriptive statistics of estimated fixed effect parameters are summarised in Table 3.

Table 2. M and SD of SRT by Condition for ASC group (Experiment 1).

Condition	Random (Arrows)	NoSearch (Arrows)	Predictive (Arrows)	Random (Eyes)	NoSearch (Eyes)	Predictive (Eyes)
M (SD)	474.98 (351.42)	385.17 (225.78)	343.15 (263.28)	476.17 (334.89)	445.70 (287.51)	341.16 (196.93)

Note. Means and standard deviations are provided in the format M(SD).

Table 3. Estimated fixed effect parameters for SRT (Experiment 1).

Effect	β-coefficient	Standard Error (SE)	t-ratio	p-value
Context				

NoSearch-Random	-0.163	0.095	-1.708	0.100
Predictive-NoSearch	0.623	0.109	5.723	< .001***
Predictive-Random	-1.572	0.215	-7.323	< .000***
Stimulus				
Arrows-Eyes	-0.170	0.083	-2.043	0.048*
Stimulus*Context				
NoSearch-Random	-0.241	0.070	-3.445	0.001***
Predictive-NoSearch	0.144	0.072	2.000	0.046*
Random-Predictive	-0.097	0.075	-1.294	0.196
Follow-up comparisons				
NoSearch-Random (Arrows)	-0.283	0.102	-2.782	0.009**[a]
NoSearch-Random (Eyes)	-0.042	0.101	-0.416	0.680[a]
Predictive-NoSearch (Arrows)	-0.551	0.115	-4.783	< .001***[a]
Predictive-NoSearch (Eyes)	-0.696	0.114	-6.082	< .001***[a]
Arrows-Eyes (NoSearch)	0.299	0.091	3.279	0.002**[a]
Arrows-Eyes (Random)	0.058	0.094	0.615	0.541[a]
Arrows-Eyes (Predictive)	0.154	0.095	1.626	0.109[a]

Note. *$p < .05$, ** $p < .01$, *** $p < .001$, [a] Uncorrected *p*-values. However effects remain significant after applying an FDR correction for multiple comparisons (all *p*s < 0.02).

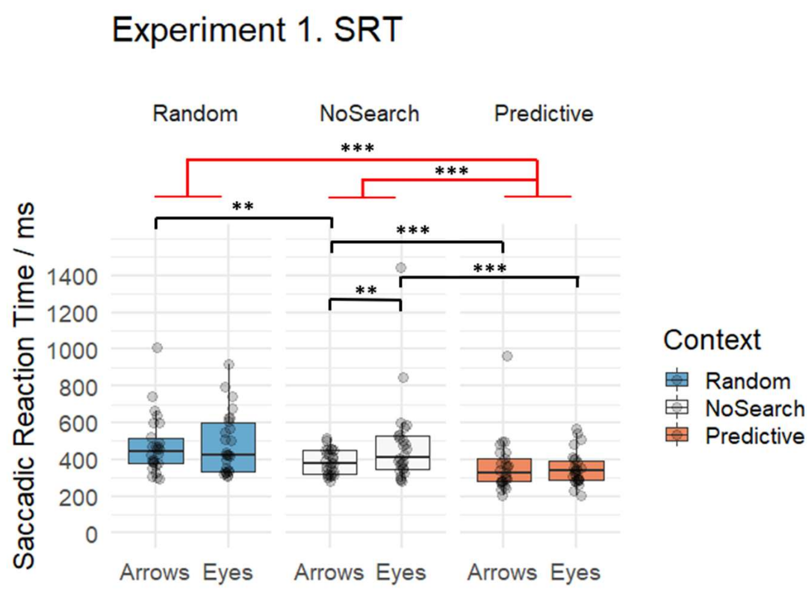

Figure 5. Boxplot with individual data points for saccadic reaction times on correct trials by context (Random Search, NoSearch, Predictive Search) and stimulus (Eyes, Arrows). Significant effects of context are illustrated in red while post-hoc contrasts are shown in black (** $p < .01$, *** $p < .001$).

Model fit analyses. For quantifying the effects of stimulus and context, model-fit-improvement was compared as a function of each fixed effect parameter. Compared to the null model (i.e., a model with no fixed-effect factors), adding the context factor significantly improved the model fit by 30.57 times ($X^2(1) = 30.57$, $p < 0.001$). Adding the stimulus factor to the context-only model improved the model fit further by 4.44 times ($X^2(1) = 4.44$, $p = 0.035$). On the other hand, including the stimulus factor to the null model first enhanced the model's fit by only 5.03 times ($X^2(1) = 5.03$, $p = .025$), while adding the context effect to the stimulus-only model significantly improved the model fit 29.97 times ($X^2(1) = 29.97$, $p < .001$). Critically, compared to a model containing fixed-effect factors for both stimulus and context, adding the interaction parameter significantly improved the model fit by 12.12 times ($X^2(1) = 12.12$, $p = .002$). These analyses show a larger effect of context than stimulus.

However, it also suggests that both factors explain unique variance in the data and that the data are best explained by a model that specifies a stimulus-by-context interaction.

2.4 Discussion

In this study, participants completed an improved version of the interactive paradigm from Caruana et al. (in prep.) where we implemented a better-matched control condition to investigate the effect of context on joint attention ability in neurotypical adults. Participants' accuracy and SRT data were analysed in three context (Random Search, NoSearch and Predictive Search) and two stimulus (Eyes and Arrows) conditions. Overall, participants made more errors in the Random context than both the NoSearch and Predictive Search contexts. There was no overall significant difference between Predictive and NoSearch contexts on accuracy. There was no overall accuracy differences between Eyes and Arrows nor a stimulus-by-context interaction when comparing the Random or Predictive Search contexts to NoSearch. There was, however, a significant interaction when comparing the Random to the Predictive Search context. This was characterised by a larger difference between contexts for Arrows than Eyes. These findings indicate that establishing joint attention was more difficult in the Random context than both the NoSearch and Predictive search contexts. Results also confirm earlier findings of a relative advantage for Eyes than Arrows when comparing between contexts that were equally complex but differed in the relevance of spatial information conveyed (i.e., Random and Predictive Search). However, given the evident ceiling effect for accuracy in the NoSearch and Predictive Search contexts (see Figure 4), these accuracy results need to be interpreted with caution. Future studies using a similar paradigm with a more complex task would be useful to better understand the effect of context on the ability to identify accurately and respond to communicative joint attention bids.

The saccadic reaction time results indicate that, in line with our previous findings, participants were faster to respond to Arrows than Eyes on the NoSearch context (Caruana et al., in prep.; Caruana, McArthur, Woolgar, & Brock, 2017a). Based on the ostensive-inferential communication model (Sperber & Wilson, 1986; Wilson & Sperber, 2002), ostensive signals (e.g., eye contact) are hypothesised to have the unique ability to activate mentalising processes that help infer the communicative intent of a social partner. This activation of higher-order social cognitive and mentalising processes may lead to an increase in cognitive processing load. This could thus result in a delayed response to Eyes compared to Arrows. This interpretation is in line with findings indicating that direct gaze results in a rapid and automatic activation of subcortical pathways associated with the social brain network (Conty, N'Diaye, Tijus, & George, 2007; Mares, Smith, Johnson, & Senju, 2016; Senju & Johnson, 2009b).

Alternatively, another possible explanation for this effect could be that the Arrows stimuli used were perceptually more salient than the Eyes. This is because our Arrows and Eyes stimuli were not perfectly matched on low-level visual properties (e.g., luminance, number of pixels, etc.). Such differences, therefore, may have resulted in faster responses to the more salient Arrows stimuli because they were perceived more easily. This is a challenge in social cognition research using eye-gaze stimuli because it makes it more challenging to identify whether the effects reported are due to the social nature of the social stimuli, or due to the perceptual salience of non-social stimuli. Future studies matching for the low-level visual properties between the Arrows and Eyes are needed to investigate the influence of the salience of stimuli to verify this effect.

The stimulus effect in the Predictive Search context and the stimulus-by-context interaction when comparing the Random and Predictive contexts did not reach significance. These effects were previously found to be significant, with evidence for faster responses on

Predictive Arrows than Predictive Eyes, and a larger effect of context in Arrows than Eyes (Caruana et al., in prep.). These unreplicated effects could be explained by the better-matched Arrows condition, where a yellow fixation point was used to match the direct gaze stimuli. Previously a green fixation point was used to match the direct gaze stimuli. This fixation point also visually matched the green arrow stimuli presented prior to or after the fixation point. This could have resulted in a less salient stimulus change in the Arrows when compared to the direct and averted gaze stimuli in the Eyes condition. Therefore, in this study a yellow fixation point was used to provide a more salient stimulus that matched the effect of eye contact in capturing attention. This added salience of the fixation point could have affected the attention switching ability by slowing responsivity, making the performance in Arrows more comparable to that of the Eyes. Hence, this suggests that we were successful in making the conditions more perceptually matched.

Importantly in this study, we were able to compare both the Predictive and Random contexts to a ‘context-free’ NoSearch baseline condition. The inclusion of this condition allowed us to decipher the overall effects attributable to contextual information (i.e., predictive or random spatial information) on joint attention responsivity. We were also able to compare the ability to identify relevant cues in the social (Eyes) domain to a matched non-social (Arrows) condition to determine whether effects were specific to the social domain. Compared to the NoSearch condition, there was an overall advantage for the Predictive context and no significant overall difference for the Random context. Also, there was an evident relative advantage for Eyes in both Random and Predictive contexts. This was apparent in the larger advantageous effect of predictive spatial sequence on response times in Eyes than Arrows. Also, there was a detrimental effect associated with a random spatial sequence on response times only with Arrows and not Eyes. As expected, this confirms our previous findings of a relative advantage for evaluating relevant Eyes stimuli within context compared to Arrows stimuli. This evident social advantage for evaluating contextual information and selecting the relevant information could be due to two factors: (1) the inherent ability to infer communicative intent from ostensive signals, which is unique to social interactions and creates an expectation for a relevant informative cue, or (2) the higher ecological validity of the social (Eyes) condition when compared to the non-social (Arrows) condition.

The ostensive-inferential communication model proposes that we are predisposed to attend to ostensive signals (e.g., eye contact) because we expect them to convey others' intention to communicate a relevant information to us (Wilson & Sperber, 2002). Thus, our sensitivity to ostension could possibly support our ability to detect relevant information to successfully guide social interactions. This claim has been investigated and supported by evidence suggesting that ostensive signals used in conveying communicative intent (e.g., eye contact), resulted in a stronger activation of brain regions associated with mentalising processes compared to non-ostensive, non-communicative social stimuli (e.g., averted gaze; Kampe, Frith, & Frith, 2003). A number of studies have attributed this effect to the detection of communicative intent (Conty et al., 2007; Kampe et al., 2003; Schilbach et al., 2006). Therefore, this unique ability could justify the relative advantage for Eyes reported by our study.

Alternatively, this effect could also be due to the social condition being more ecologically valid than the non-social condition. To achieve an optimal level of control we had to design an arrow stimulus that 'behaves' in a rather social manner by being ambiguous and requiring to be evaluated for relevance. This behaviour is not typically expected from non-social arrow stimuli in real-life, since they are always used to convey information. Therefore, this resulted in a counterintuitive need for evaluating the informative value (i.e. relevance) of a non-social referential cue (i.e. an arrow), which is typically experienced as a static informative cue and is rarely, if ever, non-informative. This non-realistic behaviour may have resulted in the decreased responsivity in the non-social Arrows condition. This is problematic as it makes it hard to conclude which factor drives the effects reported. It is currently not clear if the effects reported were due to a socially-specific advantage for ostensive signals, or due to simply having more experience with ambiguous social information than ambiguous non-social information. As indicated in our previous work, this raises questions regarding the effect of ecological validity of non-social control stimuli in studies of joint attention and social interaction in general (Caruana et al., in prep.). It is important for future work to attempt to address this issue to help in designing better control conditions using second-person paradigms.

Chapter 3: Effect of context in autism (Experiment 2)

3.1 Introduction

This experiment sought to understand the ability of autistic adolescents and young adults to identify communicative (i.e., relevant) gaze cues when embedded in a realistic context of non-communicative (i.e., irrelevant) gaze behaviour. We were interested in studying the effect of this ability on joint attention responsivity. To do this, we utilised the same paradigm used in Experiment 1. However, as a first step, we only compared performance across two contexts (Random and NoSearch) with both stimuli (Eyes and Arrows). This allowed for the comparison of previous findings using a similar paradigm in autism (Caruana et al., 2018) with the current findings using (1) a better matched control (Arrows) condition which is matched in the need to evaluate the relevance of guiding cues, (2) a context-free (NoSearch) baseline condition. Hence, we were able to verify if the previously-reported decreased responsivity with the Eyes stimuli was indeed specific to the social domain or if it was due to the unmatched need for contextual processing on the Arrows control condition. We were also able to investigate the effect of contextual processing by comparing with a 'context-free' condition. Autistic participants' responses were compared to a comparison group of neurotypical adolescents and young adults. We analysed participants' ability to respond accurately to their partner's communicative cues by looking in the correct direction to achieve joint attention successfully. We also analysed their saccadic reaction times (SRTs) to communicative gaze cues on accurate trials.

3.1.1 Hypotheses

Based on previous findings in autism (Caruana et al., 2018), and consistent with the ostensive-inferential communication model (Sperber & Wilson, 1986; Wilson & Sperber, 2002), we hypothesised that autistic participants would show decreased responsivity when communicative gaze-cues were embedded in a realistic context of non-communicative gaze shifts. Therefore, we expected to find the following:

Accuracy. Participants in both groups would make more errors on the Random Search context than the NoSearch context. We also expected, however, that autistic participants would make more errors compared to typical participants with the Eyes stimuli rather than Arrows. Finally, we expected that the magnitude of the difference between Eyes and Arrows will be larger for autistic individuals than neurotypicals.

SRT. Participants in both groups would be slower to respond on the Random Search than the NoSearch conditions. We hypothesised that participants in the autistic group will be slower to respond on the Random Eyes but not Random Arrows when compared to the typical group. We expected the typical group to show a relative advantage for identifying relevant gaze cues than relevant arrow cues when comparing the Random Search to NoSearch. This prediction is consistent with results previously reported in adults (Caruana et al., in prep.). However, we expected that this relative advantage would not be replicated in the autistic group. We did not expect to find significant differences between groups on the NoSearch context.

3.2 Methods

3.2.1 Ethics Statement

All procedures implemented in this study were approved by the Macquarie University Human Research Ethics Committee (ID: **3775**). All participants gave written, informed consent to take part in this study prior to participation. Parental consent for younger participants (<18 years) was also obtained.

3.2.2 Participants

Thirty-two adolescents and young adults participated in the study, including 16 autistic participants and 16 typical participants. All participants reported normal or corrected-to-normal vision and. Four participants in the autistic group reported having a co-occurring ADHD diagnosis. No typical participants had a history of neurological injury or impairment. The autistic and typical groups were matched for age, and gender, and intellectual ability, as measured by the Wechsler Abbreviated Scale of Intelligence – Second Edition (WASI-II) (see Table 4). We used the Autism Diagnostic Observation Schedule 2 – Module 4 (Lord et al., 2012), as well as the Social Communication Questionnaire (SCQ - Lifetime; Rutter, Bailey, & Lord, 2003) to determine where they lie on the autism spectrum. The SCQ – Lifetime was also used to confirm that participants in the typical group did not have elevated autistic features (all scored well below the cut-off score for autism of 15; Rutter et al., 2003). Participant demographics and measures are summarised in Table 4.

Table 4. Participant demographics and measures.

	Autistic (n=16)	Typical (n=16)	t	df	*p-value*
Age (years)	15.9 (4.1) [22-10]	15.6 (3.9) [22-10]	0.176	30	0.862
Gender (male:female)	10:6	10:6	0	30	1
WASI-II	107.4 (18.1) [137-72]	114.9 (12.7) [133-85]	1.370	30	0.181
SCQ (Lifetime)	22.2 (6.1) [32-6]	1.8 (1.7) [6-0]	12.942	17[a]	<.0001***
ADOS – 2, Social + Comm	9.4 (6.2) [18-0]	-	-	-	-

Note. Means and standard deviations are provided in the format M(SD) [range]. *** $p < .001$.

[a] t-test was performed assuming unequal variance as indicated by Levene's test

Participants were recruited from the general community, and some adult typical participants were recruited via undergraduate student pools at Macquarie University. Participants were compensated with either money or university course credit. Upon data collection, one autistic participant's data from one block (Random Arrows condition) were excluded from the SRT analysis due to excessive error (~2 SD) that resulted in only two correct trials which were more than 2.5 SD away from the mean and were hence considered outliers. This participant had the lowest overall IQ of the autistic group (IQ=72, borderline) and was administered the Random Arrows block as their first block. Hence their poor performance could be attributed to poor comprehension of instructions. Since we are using a Linear Mixed Effects (LME) model for the analysis, which is unaffected by unbalanced observations across conditions (Quené & Van den Bergh, 2004, 2008), the same participant's data from other conditions were retained. We also confirmed that the pattern of effects reported below were unchanged by excluding this participant's data from the analysis entirely.

3.2.3 Stimulus and Apparatus

The same apparatus used in Experiment 1 (Chapter 2) was used for this experiment. The only difference was in the viewing distance, which was changed to 104 cm instead of 80 cm. This was due to renovation work that took place at the eye-tracking laboratory. Therefore, in this experiment, the avatar face subtended a visual angle of 4.84° x 2.97° and each house subtended 2.85°, and presented as previously shown in Figure 1. The arrow stimulus of the control condition subtended 0.86°, and the fixation point subtended 0.23°.

3.2.4 Design and procedure

Participants played the same interactive game from Experiment 1 but with only two context conditions (Random and NoSearch contexts). Participants were informed that they would be playing this game in collaboration with a member of our research team, named Alan. Similar to the paradigm described in Experiment 1 (Chapter 2), participants had to search for the burglar in houses with blue doors presented at the top of the screen, while Alan searched the houses at the bottom of the screen. Once the burglar was found, the player who found the burglar was required to guide their partner to the correct location. The other player was required to respond by looking at the cued location. Participants were not given explicit instructions as to how they should communicate with their partner. They were only told that they needed to work collaboratively to catch the burglar.

Participant's eye movements were recorded in four conditions: two *context* conditions (Random, NoSearch) and two *stimulus* conditions (Eyes and Arrows). Participants completed four blocks of trials of each context and stimuli combination, with 30 responding trials and 30 initiating trials each block. Again, the eye gaze and arrow conditions in each context were always administered consecutively (i.e., Eyes Random Search, Arrows Random Search, Eyes No Search, Arrows No Search) to minimise task-switching. Within each block, trial order was randomised to ensure that the location of the burglar and the number of gaze shifts made by the avatar were not conflated with order effects. These trial level features were also counterbalanced across conditions per block.

Context Conditions

The same Random Search and NoSearch conditions used in Experiment 1 (Chapter 2) were used for this experiment.

3.2.5 Statistical Analyses

Similar to the analysis in Experiment 1 (Chapter 2), the current analysis was exclusively focused on responding trials. We used the same protocol for processing accuracy and eye-tracking data. The raw data were screened and analysed using R and the full data set and R code with the analysis outputs and annotated descriptions can be found at the Open Science Framework (https://osf.io/7hdj8/). Statistical analyses of logistic and linear mixed random effects (LME) were conducted for accuracy and SRT respectively. The maximum likelihood estimation method was implemented in these analyses using the *lme4* R package (Bates & Sarkar, 2005), and p-values were estimated using the *lmerTest* package (Kuznetsova et al., 2015). We again included traditional ANOVA analyses in the supplementary R code and output for comparisons.

We were specifically interested in investigating the main effect of stimulus (Eyes, Arrows), context (Random Search, NoSearch), group (autistic, control) and the interaction effects between these factors. To do this we used custom contrast coding (i.e., 0.5,-0.5) defined for each one of our 2-level fixed factors. This allowed us to produce parameter estimates that were interpretable in a similar way as one would interpret the output of traditional 2x2x2 ANOVA (Protopapas, 2014), with parameter estimates for the overall effect of stimulus, context and group. Parameter estimates were also obtained for the interaction effects of context and stimulus, context and group as well as the context-by-stimulus-by-group interaction. We were also interested in testing the main effects of context and stimulus and their interaction within each group separately to determine whether each group demonstrated the same pattern of behaviour observed previously in neurotypical adults (see Chapter 2). This was not possible using the custom contrasts defined above, hence we used the emmeans package to manually define these missing comparisons (Lenth et al., 2019). Finally, we ran a post-hoc analysis for SRT to assist in interpreting significant interaction effects. An FDR correction was then applied to confirm significance after correcting for multiple comparisons (Benjamini & Hochberg, 1995). Any discrepancies have been reported in the results section bellow and detailed analyses with corrected p-values are included in the accompanying R code and output (https://osf.io/7hdj8/).

Accuracy and SRT models were defined with maximally-defined random-factor structures, including random intercepts for trial and by-subject random slopes for the intercept and two fixed effects (i.e. stimulus and context) (Barr et al., 2013). For SRT analyses, the residuals of the raw data violated the normality assumption and hence data were transformed using an inverse transformation (details can be found in accompanying R code and output (https://osf.io/7hdj8/); see Balota et al., 2013). All analyses had a significance criterion of $\alpha = .05$.

For estimating effect size, the chi-squared goodness-of-fit test was performed comparing a number of mixed random-effects models using Chi-square likelihood ratios to quantify the contribution of each fixed effect as well as the interaction parameter to the model fit (Johnston et al., 2006). See Chapter 2, section 2.2.5, for detailed description and rationale.

3.3 Results

Accuracy. In this experiment, we initially investigated whether and how the presence of random spatial signals differentially affected participants' ability to respond correctly to subsequent eye gaze and arrow cues. The accuracy data for context and stimulus conditions are illustrated by group in Figure 6. Participants made significantly more errors on the Random context than the NoSearch context. There was no significant main effect of stimulus or group. There were also no significant interactions in context-by-stimulus, context-by-group, stimulus-by-group or a three-way, stimulus-by-context-by-group interactions.

Upon plotting the accuracy data, we noticed a ceiling effect, in which participants performed at the upper threshold, on the easier NoSearch context condition. This is evident in Figure 7 below. This ceiling effect may have resulted in main effects and interactions not reaching significance, even though there was a significant effect in the independent variable on which participants did not perform at ceiling (i.e., Random Search; see Lewis-Beck, Bryman, & Liao, 2003). Therefore, a post-hoc analysis was conducted to analyse group and stimulus effects in the Random Search context. This analysis indicated that there was a significant effect of group with the Random Eyes conditions but not the Arrows. There were no significant effects of stimulus within any group in the Random condition. It is important to note, however, that this result did not survive a FDR correction for multiple comparisons. Also, since this effect was a result of an exploratory post-hoc analysis, it needs to be interpreted with caution.

Overall, in the typical group, approximately 4% of trials comprised errors, with the majority being Location errors (M = 3.02% of trials, M_{Error}= 73.42% of errors, SD = 7.28). Participants only made Timeout errors on 0.63% of trials (M_{Error}= 15.19% of errors, SD = 2.47) and Search errors on 0.39% of trials (M_{Error}= 9.49% of errors, SD = 1.85). In the autistic group, error trials made up around 12% of trials, with the majority being Timeout errors (M = 5.13% of trials, M_{Error}= 42.92% of errors, SD = 16.72), followed by Location errors (M = 3.65% of trials, M_{Error}= 30.50% of errors, SD = 7.69) and Search errors (M = 2.96% of trials, M_{Error}= 24.84% of errors, SD = 10.53). Descriptive statistics of estimated fixed effect parameters are summarised in Table 5.

Table 5. Parameter estimates for Accuracy (Experiment 2)

Fixed effect	β-coefficient	Standard Error (SE)	z-value	p-value
Context				
NoSearch-Random	-2.167	0.478	-4.530	< .001***
Stimulus				
Arrows-Eyes	-0.015	0.483	-0.031	0.975
Group				
Autistic-Typical	1.049	0.686	1.529	0.126

Interactions				
Stimulus*Context	0.254	0.483	0.527	0.599
Context*Group	-0.350	0.807	-0.433	0.665
Stimulus*Group	0.647	0.895	0.723	0.470
Stimulus*Context*Group	1.185	0.860	1.377	0.168
Follow-up comparisons				
Autistic-Typical (Random, Eyes)	-1.494	0.712	-2.099	0.036*[a]
Autistic-Typical (Random, Arrows)	-0.255	0.849	-0.300	0.764[a]
Arrows-Eyes (Random, Autistic)	0.507	0.626	0.811	0.418[a]
Arrows-Eyes (Random, Control)	-0.732	0.670	-1.092	0.275[a]

Note. *$p < .05$, ** $p < .01$, *** $p < .001$, [a] Uncorrected *p*-values. Significant effect did not survive a FDR correction for multiple comparisons.

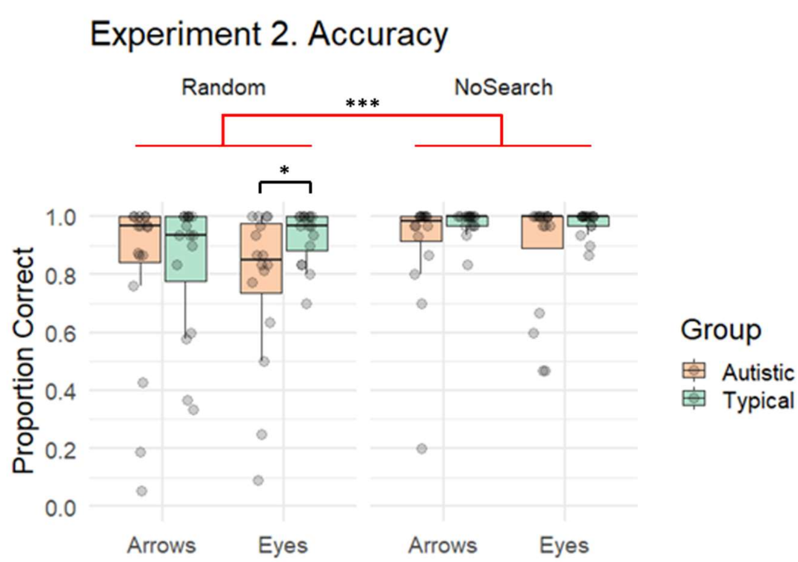

Figure 6. Boxplot with individual data points illustrating the proportion of correct trials by context (NoSearch, Random Search), stimulus (Eyes, Arrows) and Group (Autistic, Typical). Significant effect of context is illustrated in red while post-hoc contrasts are shown in black (* $p < .05$, *** $p < .001$).

SRT. We then investigated whether and how the presence of non-predictive spatial signals differentially influenced the speed with which participants in both groups were able to initiate saccadic eye movements in response to eye gaze and arrow cues. The SRT data for context and stimulus conditions are illustrated by group in Figure 7. Results showed a main effect of stimulus where participants were significantly faster in responding with Arrows stimuli than Eyes. There also was a main effect of context where participants were significantly slower when responding on Random than NoSearch context. However, there was no significant difference overall between groups. There was also a significant stimulus-by-context interaction, as well as a significant context-by-group interaction. There were no significant stimulus-by-group or stimulus-by-context-by-group interactions.

We used emmeans to determine the main effects and interactions within each group separately. For the autistic group, there was a significant effect of stimulus where autistic participants were overall slower for Eyes than Arrows. However, there was no significant effect of context or a context-by-stimulus interaction in this group. In the typical group, the effects were reversed, with a significant effect of context indicating faster responding on NoSearch than Random Search conditions. There also was a significant context-by-stimulus interaction. There was no significant effect of stimulus overall in this group.

We also used emmeans to conduct post-hoc pairwise comparisons of interest using custom contrasts, and applied a FDR correction for multiple comparisons. These comparisons confirmed that there was no significant between-group differences. There also were no significant within-group differences in the autistic group. For the typical group, there was a significant stimulus effect in the NoSearch context showing a slower response for Eyes than Arrows. This effect was not found in the Random context. The context effect indicates that participants were significantly slower in the Random context only with Arrows but not Eyes.

Mean SRT data are summarised by condition for each group in Tables 6. Descriptive statistics of estimated fixed effect parameters are summarised in Table 7.

Table 6. Saccadic Reaction Time M and SD by Condition for each group (Experiment 2).

Condition	NoSearch (Arrow)	Random (Arrow)	NoSearch (Eyes)	Random (Eyes)
Autistic M(SD)	454.67 (278.00)	536.99 (433.73)	498.33 (340.94)	510.41 (387.62)
Typical M(SD)	368.92 (173.92)	520.83 (368.93)	418.27 (225.53)	531.46 (411.59)

Note. Means and standard deviations are provided in the format M(SD).

Table 7. Parameter estimates for SRT (Experiment 2)

Fixed effect	β-coefficient	Standard Error (SE)	t-ratio	p-value
Context				
NoSearch-Random	0.265	0.065	4.088	< .0001***
Stimulus				
Arrows-Eyes	0.185	0.063	2.927	0.006**
Group				
Autistic-Typical	-0.182	0.182	-1.005	0.322
Interactions				
Stimulus*Context	0.184	0.064	2.885	0.004**
Context*Group	-0.304	0.130	-2.346	0.026*
Stimulus*Group	0.031	0.117	0.266	0.792
Stimulus*Context*Group	-0.140	0.127	-1.096	0.273
Within group effects				
Context				
NoSearch-Random (Autistic)	-0.226	0.194	-1.162	0.253
NoSearch-Random (Typical)	-0.834	0.186	-4.475	< .0001***

Stimulus				
Arrows-Eyes (Autistic)	0.401	0.181	2.223	0.032*
Arrows-Eyes (Typical)	0.340	0.175	1.940	0.060
Stimulus*Context				
Autistic	-0.114	0.093	-1.231	0.218
Typical	-0.254	0.088	-2.876	0.004**
Follow-up comparisons				
Autistic-Typical (NoSearch, Arrows)	-0.354	0.221	-1.604	0.118[a]
Autistic-Typical (NoSearch, Eyes)	-0.315	0.215	-1.467	0.151[a]
Autistic-Typical (Random, Arrows)	0.020	0.220	0.091	0.928[a]
Autistic-Typical (Random, Eyes)	-0.081	0.185	-0.437	0.664[a]
Arrows-Eyes (NoSearch, Autistic)	0.258	0.099	2.618	0.011*[a]
Arrows-Eyes (Random, Autistic)	0.144	0.105	1.377	0.173[a]
NoSearch-Random (Arrows, Autistic)	0.170	0.108	1.572	0.122[a]
NoSearch-Random (Eyes, Autistic)	0.056	0.107	0.522	0.604[a]
Arrows-Eyes (NoSearch, Typical)	0.297	0.095	3.124	0.003**[a]
Arrows-Eyes (Random, Typical)	0.043	0.101	0.427	0.671[a]
NoSearch-Random (Arrows, Typical)	0.544	0.105	5.178	< .0001***[a]
NoSearch-Random (Eyes, Typical)	0.290	0.101	2.871	0.006**[a]

Note. *$p < .05$, ** $p < .01$, *** $p < .001$, [a] Uncorrected p-values. However effects remain significant after applying an FDR correction for multiple comparisons (all ps < 0.03)

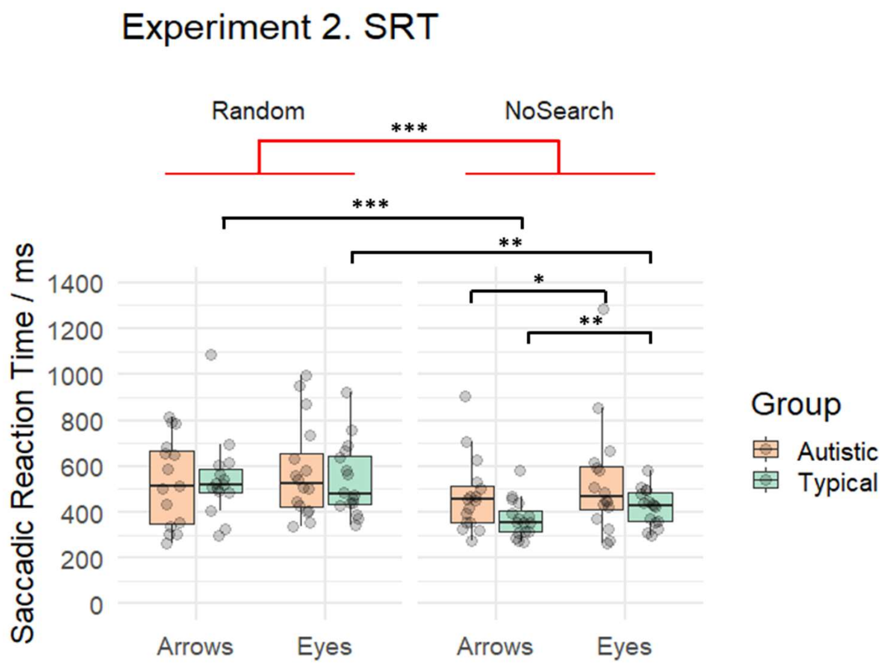

Figure 7. Boxplot with individual data points for saccadic reaction times on correct trials by context (NoSearch, Random Search), stimulus (Eyes, Arrows) and Group (Autistic, Typical). Significant effect of context is illustrated in red while post-hoc contrasts are shown in black (** p < .05, ** p < .01, *** p < .001).

Model fit analyses. For quantifying the effects of stimulus, context and group, model-fit-improvement was compared as a function of each fixed effect parameter. Compared to the null model (i.e., a model with no fixed-effect factors), adding the context factor improved the model fit by 15.03 times ($X^2(1) = 15.03$, $p < 0.001$). Adding the stimulus and group factors to the context-only model improved the model fit further by 9.41 times ($X^2(1) = 9.41$, $p = 0.009$). On the other hand, including the stimulus factor to the null model first enhanced the model's fit by 12.74 times ($X^2(1) = 12.74$, $p < .001$), while adding the context and group effects to the stimulus-only model improved the model's fit 11.70 times ($X^2(1) = 11.70$, $p = .003$). However, adding the group factor to the null model first did not have a significant effect on the model's fit ($X^2(1) = 0.26$, $p = .611$) and adding the context and stimulus effects to the group-only model improved the fit 24.18 times ($X^2(1) = 24.18$, $p <$ .0001). Critically, comparing a model containing fixed-effect factors for stimulus, context and group, to a model including the interaction parameter significantly improved the model fit by 14.79 times ($X^2(1) = 14.79$, $p = .005$). These analyses show an equivalent context and stimulus effect with context being slightly larger and no main effect of group. However, it also suggests that these factors explain unique variance in the data, and that the data are best explained by a model that specifies a stimulus-by-context-by-group interaction.

3.4 Discussion

This experiment provided the first objective assessment of joint attention in autistic youth using an ecologically valid paradigm. In this study, we sought to investigate the ability of autistic adolescents and young adults to identify and respond to communicative gaze cues when embedded in a realistic context. This enabled us to elucidate the aspects of joint attention that lead to difficulty in responsivity in autism. To do this, participants played the 'Catch the Burglar' game similar to that used in Experiment 1. Participants' accuracy and SRT data were analysed in two contexts (Random Search and NoSearch) and two stimulus (Eyes and Arrows) conditions. Overall, all participants made more errors in the Random context than the NoSearch context. There were no significant differences between the autistic and typical groups on accuracy. Given the evident ceiling effect in the objectively easier NoSearch context (see Figure 7), however, post-hoc analyses were run to check for between- and within-group differences in the more challenging Random context. This analysis revealed that autistic participants made significantly more errors on the Random Eyes condition than typical participants. This result suggests that autistic participants found it more difficult than typical participants to respond to joint attention bids within the more realistic (i.e., Random Search) context. This effect, however, was not found on the non-social (Arrows) condition, which suggests a specifically social challenge in autism.

According to the ostensive-inferential communication model (Wilson & Sperber, 2002), the ability to identify ostensive signals enables neurotypical people to infer their partner's communicative intention and identify guiding cues more accurately. This effect was not found in the autistic group where evaluating eye-gaze information was found to be more challenging for autistic than neurotypical participants. Therefore, this could indicate that eye contact does not have the same advantage in being evaluated as an ostensive signal in autism as it does in typical development. That said, it is important to note that the accuracy results reported in our experiment need to be considered with caution, as they were performed as a post-hoc analysis due to the ceiling effect in the 'NoSearch' condition. This effect also did not survive FDR correction for multiple comparisons. Future studies using a similar paradigm with a more challenging task are needed to prospectively test and confirm these findings.

The saccadic reaction time analysis revealed no significant group difference in both contexts. There was, however, a significant interaction between stimulus and context. This was verified by post-hoc pairwise comparisons revealing that participants were faster to respond to Arrows than Eyes in the NoSearch but not the Random search context in both groups. This suggests that there is a fundamental advantage for processing and responding to non-social cues presented in unambiguous contexts. This finding is in line with our previous findings in neurotypical adults, where this effect was attributed to the activation of higher-order social cognitive processes (Experiment 1 (Chapter2); Caruana et al., in prep.; Caruana et al., 2017a). This activation, which is specific to the social (i.e., Eyes) condition, may have resulted in the slower responsivity reported. It is also possible, however, for this effect to be due to the Arrows stimuli being perceptually more salient than the Eyes. As noted in section 2.4, our stimuli were not matched on low-level visual properties (e.g., luminance, number of pixels, etc.). It is, therefore, possible that the faster responsivity for Arrows was due to the stimuli being easier to perceive than Eyes. This could be especially important for autistic participants as there is some indication of atypical low-level visuoperceptual processing in autism (Milne et al., 2002). There are also, however, contradictory findings suggesting no low-level processing differences in autism (e.g., Pellicano, Gibson, Maybery, Durkin, & Badcock, 2005). Future studies matching for the low-level visual salience of eye and arrow stimuli are needed to verify these stimulus effects on responsivity and to confirm whether these effects are driven by the conceptual stimulus category or low level properties.

There also was a significant interaction between context and group indicating that the effect of context was larger in the typical than in the autistic group. Follow-up comparisons revealed a significant effect of context in the typical group only, suggesting that typical adolescents and adults found it easier to process and respond to cues presented in an unambiguous context (NoSearch) than one which required them to evaluate the relevance of spatial cues before responding (Random search). Autistic participants, however, did not significantly differ in responsivity across contexts, but at the same time, they were also not significantly slower than typical participants.

Finally, there also was a significant stimulus by context interaction in the control group indicating a larger effect of context with the Arrows than the Eyes. This suggests that the cost of embedding Arrows in context was generally larger than the cost of embedding Eyes in context for the typical group. This finding indicates a socially-specific advantage for evaluating gaze-cues in the realistic context, similar to our previous findings with neurotypical adults as discussed in Experiment 1. These results suggest that typically developing adolescents and young adults identify eye contact as an ostensive signal and utilise it to infer the communicative intent of a social partner. This, in turn, helps them correctly infer the informative value of the following guiding gaze cue. Therefore, the unique ability for ostensive cues to elicit the communicative intent of a partner is thought to be responsible for the relative advantage in identifying relevant eye-gaze cues by neurotypical people. However, of interest here, these effects were not replicated in the autistic group, indicating no evidence for autistic participants having a specific advantage for Eyes during the realistic and complex joint attention context.

Our results from the autistic group are inconsistent with our initial hypothesis for the Random Search in the social (i.e., Eyes) condition, and previous findings in autistic adults (Caruana et al., 2018). Based on previous findings, we expected to find a slower responsivity to gaze-cued joint attention in the autistic compared to the typical group. Our findings, however, did not reveal any significant difference in SRT between groups. This could be due to the difference in design between both studies. In the previous experiment (Caruana et al., 2018), autistic adults performed two blocks of the Random Search condition by two stimuli, with each block comprising 27 trials. Participants were found to be slower to respond on the social Eyes condition in the first block but not the second block. It could be that this effect was not replicated in our current study due to the increased number of trials per block (60 trials). This could have resulted in the overall effect not reaching significance even though the learning pattern could be different. It is possible that autistic participants show a significant difference only on the earlier trials. This difference could be quickly decreasing as the trials proceed and participants learn how to perform the task. However, we were unable to verify whether such a learning effect occurred using the current data set given that the current study included several stimulus and context conditions that were administered in a counterbalanced order. Therefore, learning (i.e., a reduction in responsivity speed) may have occurred as a function of block order and context/stimulus exposure. For instance, it could be that performance on the second block was always better than the first due to prior task exposure. As such, a future study with a larger sample that is only exposed to the Random condition is needed to investigate whether learning occurs during this task specifically – and whether the learning rate differs in autism.

Our findings on the non-social Arrows condition, however, were consistent with our initial hypothesis of no significant difference between groups on the Random Search. Therefore, our study did not reveal any evidence for global difficulties in evaluating contextual clues in autistic people. However, our findings do reveal that unlike typical young people, young autistic people do not exhibit a unique advantage for Eyes compared to Arrows in our ‘context-dependent’ task (i.e., Random Search). This suggests a lack of sensitivity to ostensive signals. Unlike our hypothesis, however, when an ostensive signal is correctly evaluated, autistic participants responded to the following referential cue appropriately and their response time was not different to that of the typical group. This finding aligns with evidence suggesting that autistic individuals show a reduced sensitivity, but not an inability, to identify ostensive eye contact cues, and use them less when evaluating the relevance of upcoming gaze shifts (Böckler et al., 2014; Senju & Johnson, 2009a). Other evidence also suggest that autistic children do not show the same preferential detection for direct gaze compared to averted gaze as that evident in neurotypical people (Senju, Yaguchi, Tojo, & Hasegawa, 2003).

There is compelling evidence to suggest that the active participation of infants in social learning opportunities provides an important foundation for both cognitive and communicative development in typical development (Mundy & Neal, 2000; Tomasello et al., 1993; Ulvund & Smith, 1996). The diminished eye contact effect and the reduced sensitivity to ostensive signals in autism could result in a compromised opportunity to learn from others. The lack of sensitivity for ostensive signals in autism has, therefore, been hypothesised to be related to the deficits exhibited by the condition in joint attention, ToM and social learning (Csibra & Gergely, 2009; Franchini et al., 2017; Happé, 1993; Mundy, 2016; Mundy & Neal, 2000; Sodian et al., 2015). There has been evidence, however, suggesting that autistic individuals have difficulty processing gaze and facial expressions when the stimuli were only briefly presented (Clark, Winkielman, & McIntosh, 2008; Wallace, Coleman, Pascalis, & Bailey, 2006). Our current study specifically manipulated the context in which relevant communicative gaze-cues were presented. We did not, however, manipulate the presence or duration of eye contact. Therefore, future studies need to systematically manipulate these factors to characterise and confirm the role of eye contact during responsive joint attention as an ostensive communicative cue.

Following the relevance-theoretic account of social communication (Sperber & Wilson, 1986; Wilson & Sperber, 2002), for an ostensive signal to have optimal relevance it needs to be: (1) relevant enough to the receiver for it to be worth processing; and (2) the most relevant stimulus that can be produced given the communicator's abilities and preferences. It is possible that processing eye gaze is found to be cognitively demanding for autistic individuals and is, hence, not considered relevant enough to be processed and is not evaluated as an ostensive signal as readily as in neurotypical individuals. This view is supported by evidence indicating an atypical increased subcortical activation with perceived eye contact in autism (Dalton et al., 2005; Hadjikhani et al., 2017). More work is needed to determine the importance of eye contact as an ostensive signal in the presence of other more salient signals (e.g. pointing, gesturing, directed speech) in a realistic social interaction to better understand and accommodate for the abilities and preferences of autistic people.

Finally, this study only considered investigating the ability to infer the communicative intent from an ostensive signal followed by a guiding cue when embedded in a random context of eye-gaze behaviour in autism. Future directions should investigate the effect of a predictive context (as in Experiment 1) on the ability to infer the direction of gaze-cue of a social partner in autism. This will help us further understand how those contextual information embedded in the sequence of non-communicative eye movements are used by autistic people to support gaze responsivity during joint attention. Investigating the specifically challenging aspects of responsive joint attention throughout development will assist in formulating empirically-informed learning paradigms to promote effective and supportive learning environments for autistic individuals and those who interact with them.

Summary

The current thesis was interested in investigating the ability to identify relevant communicative gaze cues when embedded in a realistic context of irrelevant, non-communicative gaze behaviour and its influence on joint attention responsivity. We achieved this by using an interactive and ecologically valid joint attention paradigm in a series of eye-tracking studies. This paradigm was adapted from Caruana et al. (in prep.), implementing a small manipulation (i.e., using a yellow fixation point instead of green) to the control condition to afford a higher degree of experimental control. Using this paradigm, we wanted to specifically test the influence of informative (i.e., Predictive) and non-informative (i.e., Random) contextual clues on joint attention responsivity in neurotypical adults (Experiment 1, Chapter 2). We also took the first step towards understanding the influence of context on joint attention responsivity in autistic adolescents and young adults by testing the influence of a non-informative (i.e., Random) context (Experiment 2, Chapter 3).

The first aim of this project was to investigate how different contextual clues affect responsivity to joint attention bids in neurotypicals. We investigated the effect of adding a (1) Random context, in which joint attention gaze cues were preceded by a sequence of non-informative and non-communicative gaze-shifts, or (2) Predictive context, in which joint attention gaze cues were preceded by an informative sequence of non-communicative gaze-shifts that was predictive of the target location. These Random and Predictive gaze contexts were compared to a context-free (NoSearch) condition, in which gaze cues were not preceded by any non-communictive eye movements. Performance on all three context conditions was also compared to matched non-social conditions in which the gaze stimuli were replaced by an arrow stimulus. Comparing the Eyes and Arrows conditions enabled us to assess whether any context effects were specific to the social domain, or rather, reflected domain-general effects on attention and/or executive function.

This work was motivated by the ostensive-inferential communication model (Sperber & Wilson, 1986; Wilson & Sperber, 2002) which makes several key claims. First, the model suggests that communicators typically express their communicative intent by utilising specific social cues known as 'ostensive signals' (e.g., eye contact). These ostensive signals are claimed to have the unique ability to capture and orient attention to relevant information. Therefore, we hypothesised that this ability would result in a uniquely social (i.e., eye gaze) advantage when responding to relevant communicative cues (i.e., guiding gaze shift) and differentiating them from irrelevant, non-communicative information (i.e., searching gaze shift) in our social joint attention task.

To this end, **Chapter 2** (Experiment 1) evaluated 26 neurotypical adults on a virtual reality joint attention task comparing performance across the three context (i.e., Random, Predictive, NoSearch) and two stimulus conditions (i.e., Eyes, Arrows). We found that participants made significantly more errors on Random context trials than the other contexts. We also found that participants were significantly faster to respond to cues during the Predictive context than both Random and NoSearch. Critically, we also obtained evidence for a significant relative SRT advantage for gaze responsivity across contexts, compared to arrow responsivity. Specifically, we found that response times to eye gaze cues – whilst slower in general compared to arrows – were less-affected by random, uninformative contextual information. Further, participants were more sensitive to the presence of predictive contextual information conveyed by eyes than arrows. Together, this relative gaze advantage was attributed to the unique advantage of eye contact as an ostensive signal.

In **Chapter 3** (Experiment 2), we took the first step towards objectively investigating the specific factors that play a role in the difficulties autistic individuals experience when responding to joint attention bids. To achieve this, we again used the same paradigm implemented in Chapter 2 (Experiment 1) in a sample of young autistic individuals (n = 16) and age- and IQ-matched controls with typical development (n = 16). In this experiment, we only compared performance across two context conditions (i.e., Random, NoSearch). We again, also had participants complete both stimulus conditions (i.e., Eyes, Arrows). Our analyses did not provide any evidence for an overall difference between groups in terms of either accuracy or SRT measures of joint attention responsivity. Significantly, however, we did find evidence that more errors and slower responses were made by participants in both groups during the Random than NoSearch contexts. As in Chapter 2 (Experiment 1), we found evidence for a relative SRT advantage for responding to Eyes than Arrows. This was again characterised by a smaller effect of Random context on responsivity for Eyes than Arrows. Interestingly, however, this relative advantage was only observed in neurotypical but not autistic participants in our sample. Together, this reveals that although young autistic individuals were able to complete all task conditions – with performance that was largely commensurate with their neurotypical peers – they may nevertheless be less sensitive to eye contact as an ostensive signal.

These findings serve as the first objective assessment of joint attention responsivity in autistic youth using an ecologically valid paradigm. Further, the capacity of our paradigm to manipulate the presence of contextual information provides a powerful tool for examining the specific aspects of joint attention that may contribute to the difficulty in responsivity in autism. This study, however, only investigated the effect of a non-informative context (i.e., Random) in autism. Future studies should attempt to investigate how informative contextual information (i.e., Predictive context from Experiment 1) might influence joint attention responsivity. It would be particularly interesting to investigate if autistic individuals utilise contextual clues to predict subsequent communicative gaze-cues as seen demonstrated by neurotypical participants in Experiment 1. Additionally, the context manipulation afforded by our paradigm provides the field with a powerful tool which can be used to empirically test the influential Bayesian inference and predictive coding accounts of autism (Pellicano & Burr, 2012; Van Boxtel & Lu, 2013; Van de Cruys et al., 2014).

Recent accounts of autism have suggested that the social and non-social difficulties characteristic of the condition are related to associating lower confidence to prior knowledge when processing current sensory input (Pellicano & Burr, 2012; Van Boxtel & Lu, 2013; Van de Cruys et al., 2014). This hypothesis is based on the principles of Bayesian inference and the predictive coding theoretical accounts of cognition (Clark, 2013; Friston & Kiebel, 2009; Kersten, Mamassian, & Yuille, 2004; Knill & Pouget, 2004). These frameworks propose that the human brain constantly evaluates and updates information and predictions about sensory events in the environment to minimise prediction errors based on incoming input. As such, the amount of confidence we place on our prior knowledge determines our propensity to update our prior beliefs. It has, therefore, been hypothesised that autistic people have a more accurate perception of the world due to associating less importance to prior experience when evaluating new sensory information (Lawson, Mathys, & Rees, 2017; Pellicano & Burr, 2012). However, this theory could also explain many of the characteristic features of autism including the potential lack of sensitivity to eye contact as an experientially identified ostensive signal. Therefore, it is possible that the optimal relevance which is claimed to be typically associated with ostensive signals, is not attributed similarly in autism. If this was the case, we would expect to find a similar decreased sensitivity to ostensive signals other than eye contact (e.g., pointing, eyebrow flash, calling ones' name). Future studies investigating this hypothesis using a similar controlled interactive paradigms will be needed to verify this claim.

Challenges faced by autistic people, however, could be related to issues much broader than a specific difficulty in cognition. Rather problems of social exclusion could be playing an equally important role due to deficits in social learning opportunities (Sasson et al., 2017). Indeed, as noted previously, many theoretical accounts have emphasised the importance of social interactions in the development and formation of higher cognitive processes (e.g., mentalising, language) by facilitating learning experiences (e.g., Csibra & Gergely, 2009; Vygotsky, 1980). It is possible, therefore, that the challenges faced by autistic individuals are aggravated by the cultural social exclusion they face due to a misalignment between the preferences of interacting social partners (Bolis, Balsters, Wenderoth, Becchio, & Schilbach, 2017; Bolis & Schilbach, 2018; Milton, 2012). It is therefore important for future work to investigate this hypothesis and possibly help dissociate between the cultural and biological aspects of the condition. Future studies should also focus on identifying the unique preferences of autistic individuals to help accommodate them within supportive social and technological environments.

References

Adamson, L. B., Bakeman, R., Suma, K., & Robins, D. L. (2019). An Expanded View of Joint Attention: Skill, Engagement, and Language in Typical Development and Autism. *Child development, 90*(1), e1. doi:10.1111/cdev.12973

Adrien, J. L., Lenoir, P., Martineau, J., Perrot, A., Hameury, L., Larmande, C., & Sauvage, D. (1993). Blind ratings of early symptoms of autism based upon family home movies. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*(3), 617-626.

Akhtar, N., Dunham, F., & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of child language, 18*(1), 41-49.

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

Argyle, M., & Cook, M. (1976). Gaze and mutual gaze.

Baez, S., García, A. M., & Ibanez, A. (2016). The social context network model in psychiatric and neurological diseases. In *Social Behavior from Rodents to Humans* (pp. 379-396): Springer.

Baez, S., & Ibanez, A. (2014). The effects of context processing on social cognition impairments in adults with Asperger's syndrome. *Frontiers in neuroscience, 8*, 270.

Baez, S., Rattazzi, A., Gonzalez-Gadea, M. L., Torralva, T., Vigliecca, N., Decety, J., . . . Ibanez, A. (2012). Integrating intention and context: assessing social cognition in adults with Asperger syndrome. *Frontiers in Human Neuroscience, 6*, 302.

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(5), 1563.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Bates, D. M., & Sarkar, D. (2005). The lme4 library.[On-line]. *Available: lib. stat. cmu. edu/R/CRAN*.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental science, 8*(6), 492-499.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

Birmingham, E., Ristic, J., & Kingstone, A. (2012). Investigating social attention: A case for increasing stimulus complexity in the laboratory. *Cognitive neuroscience, development, and psychopathology: Typical and atypical developmental trajectories of attention*, 251-276.

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Observing shared attention modulates gaze following. *Cognition, 120*(2), 292-298.

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.

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2017). Beyond autism: introducing the dialectical misattunement hypothesis and a bayesian account of intersubjectivity. *Psychopathology, 50*(6), 355-372.

Bolis, D., & Schilbach, L. (2018). Observing and participating in social interactions: action perception and action control across the autistic spectrum. *Developmental cognitive neuroscience, 29*, 168-175.

Bottema-Beutel, K. (2016). Associations between joint attention and language in autism spectrum disorder and typical development: A systematic review and meta-regression analysis. *Autism Research, 9*(10), 1021-1035.

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental retardation and developmental disabilities research reviews, 10*(3), 169-175.

Bruner, J. S. (1974). From communication to language—A psychological perspective. *Cognition, 3*(3), 255-287.

Cabeza, R., Prince, S. E., Daselaar, S. M., Greenberg, D. L., Budde, M., Dolcos, F., . . . Rubin, D. C. (2004). Brain activity during episodic retrieval of autobiographical and laboratory events: an fMRI study using a novel photo paradigm. *Journal of cognitive neuroscience, 16*(9), 1583-1594.

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience*, 159-181.

Carpenter, R. H. (1988). *Movements of the Eyes, 2nd Rev*: Pion Limited.

Caruana, N., Alhasan, A., Wagner, K., Kaplan, D., Woolgar, A., & McArthur, G. (in prep.). The effect of non-communicative eye movements on joint attention. Manuscript submitted for publication.

Caruana, N., Brock, J., & Woolgar, A. (2015). A frontotemporoparietal network common to initiating and responding to joint attention bids. *Neuroimage, 108*, 34-46.

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017a). Detecting communicative intent in a computerised test of joint attention. *PeerJ, 5*, e2899.

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017b). Simulating social interactions for the experimental investigation of joint attention. *Neuroscience & Biobehavioral Reviews, 74*, 115-125.

Caruana, N., Seymour, K., Brock, J., & Langdon, R. (2019). Responding to joint attention bids in schizophrenia: An interactive eye-tracking study. *Quarterly journal of experimental psychology*, 1747021819829718.

Caruana, N., Stieglitz Ham, H., Brock, J., Woolgar, A., Kloth, N., Palermo, R., & McArthur, G. (2018). Joint attention difficulties in autistic adults: an interactive eye-tracking study. *Autism, 22*(4), 502-512.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*(8), 1839-1849.

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage, 12*(3), 314-325.

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1430), 315-324.

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.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences, 36*(3), 181-204.

Clark, T. F., Winkielman, P., & McIntosh, D. N. (2008). Autism and the extraction of emotion from briefly presented facial expressions: stumbling at the first step of empathy. *Emotion, 8*(6), 803.

Clifford, S. M., & Dissanayake, C. (2008). The early development of joint attention in infants with autistic disorder using home video observations and parental interview. *Journal of Autism and Developmental Disorders, 38*(5), 791-805.

Conty, L., N'Diaye, K., Tijus, C., & George, N. (2007). When eye creates the contact! ERP evidence for early dissociation between direct and averted gaze motion processing. *Neuropsychologia, 45*(13), 3024-3037.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron, 58*(3), 306-324.

Couto, B., Sedeno, L., Sposato, L. A., Sigman, M., Riccio, P. M., Salles, A., . . . Ibanez, A. (2013). Insular networks for emotional processing and social cognition: comparison of two case reports with either cortical or subcortical involvement. *Cortex, 49*(5), 1420-1434.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences, 13*(4), 148-153.

Csibra, G., & Volein, A. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology, 26*(1), 1-11.

D'Entremont, B., Hains, S. M., & Muir, D. W. (1997). A demonstration of gaze following in 3-to 6-month-olds. *Infant Behavior and Development, 20*(4), 569-572.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., . . . Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature neuroscience, 8*(4), 519.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., 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:10.1037/0012-1649.40.2.271

De Jaegher, H. (2009). Social understanding through direct perception? Yes, by interacting. *Consciousness and cognition, 18*(2), 535-542.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in cognitive sciences, 14*(10), 441-447.

Delinicolas, E. K., & Young, R. L. (2007). Joint attention, language, social relating, and stereotypical behaviours in children with autistic disorder. *Autism, 11*(5), 425-436.

El Zein, M., Wyart, V., & Grezes, J. (2015). Anxiety dissociates the adaptive functions of sensory and motor response enhancements to social threats. *Elife, 4*, e10274.

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology, 85*(3), 199-212.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy, 5*(1), 39-60.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition, 57*(2), 109-128.

Franchini, M., Glaser, B., Wood de Wilde, H., Gentaz, E., Eliez, S., & Schaer, M. (2017). Social orienting and joint attention in preschoolers with autism spectrum disorders.(Research Article). *PLoS ONE, 12*(6), e0178859. doi:10.1371/journal.pone.0178859

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic bulletin & review, 5*(3), 490-495.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin, 133*(4), 694.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1211-1221.

Frith, C., & Frith, U. (2000). The physiological basis of theory of mind: functional neuroimaging studies. *Understanding other minds: Perspectives from developmental cognitive neuroscience, 2*.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron, 32*(6), 969-979.

Frith, U., & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1537), 165-176.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in cognitive sciences, 7*(2), 77-83.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind'in verbal and nonverbal tasks. *Neuropsychologia, 38*(1), 11-21.

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in cognitive sciences, 17*(4), 160-165.

Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science, 20*(3), 347-353.

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition, 136*, 359-364.

Gomez, J.-C. (2005). Joint attention and the notion of subject: Insights from apes, normal children, and children with autism.

Grice, H. P. (1957). Meaning. *The philosophical review, 66*(3), 377-388.

Hadjikhani, N., Johnels, J. Å., Zürcher, N. R., Lassalle, A., Guillon, Q., Hippolyte, L., . . . Gillberg, C. (2017). Look me in the eyes: constraining gaze in the eye-region provokes abnormally high subcortical activation in autism. *Scientific Reports, 7*(1), 3163.

Hamilton, A. F. d. C. (2016). Gazing at me: the importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1686), 20150080.

Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*(2), 101-119.

Heal, J. (2005). Joint attention and understanding the mind. *Joint attention: Communication and other minds*, 34-44.

Heatherton, T. F., Wyland, C. L., Macrae, C. N., Demos, K. E., Denny, B. T., & Kelley, W. M. (2006). Medial prefrontal activity differentiates self from close others. *Social cognitive and affective neuroscience, 1*(1), 18-25.

Hobson, J. A., & Hobson, R. P. (2007). Identification: The missing link between joint attention and imitation? *Development and psychopathology, 19*(2), 411-431.

Hobson, R. P. (2005). What Puts the Jointness into Joint Attention? *Joint attention: Communication and other minds: Issues in philosophy and psychology*, 185.

Ibañez, A., & Manes, F. (2012). Contextual social cognition and the behavioral variant of frontotemporal dementia. *Neurology, 78*(17), 1354-1362.

Jarick, M., & Kingstone, A. (2015). The duality of gaze: eyes extract and signal social information during sustained cooperative and competitive dyadic gaze. *Frontiers in Psychology, 6*, 1423.

Johnston, J. E., Berry, K. J., & Mielke Jr, P. W. (2006). Measures of effect size for chi-squared and likelihood-ratio goodness-of-fit tests. *Perceptual and motor skills, 103*(2), 412-414.

Kampe, K. K., Frith, C. D., & Frith, U. (2003). "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *Journal of Neuroscience, 23*(12), 5258-5263.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous child, 2*(3), 217-250.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol., 55*, 271-304.

Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The social attribution task. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(7), 831-846.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences, 27*(12), 712-719.

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., . . . Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function, 220*(2), 587-604.

Kuhn, G., Benson, V., Fletcher-Watson, S., Kovshoff, H., McCormick, C. A., Kirkby, J., & Leekam, S. R. (2010). Eye movements affirm: automatic overt gaze and arrow cueing for typical adults and adults with autism spectrum disorder. *Experimental Brain Research, 201*(2), 155-165.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmertest’. *R package version, 2*(0).

Kwisthout, J., Vogt, P., Haselager, P., & Dijkstra, T. (2008). Joint attention and language evolution. *Connection Science, 20*(2-3), 155-171.

Langton, S. R., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6*(5), 541-567.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature neuroscience, 20*(9), 1293.

Leekam, S. (2016). Social cognitive impairment and autism: what are we trying to explain? *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1686), 20150082.

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology, 15*(1), 77-95.

Leekam, S. R., López, B., & Moore, C. (2000). Attention and joint attention in preschool children with autism. *Developmental Psychology, 36*(2), 261.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package “emmeans”: Estimated Marginal Means, aka Least-Squares Means. *Compr. R Arch. Netw*, 1-67.

Lewis-Beck, M., Bryman, A. E., & Liao, T. F. (2003). *The Sage encyclopedia of social science research methods*: Sage Publications.

Lord, C., DiLavore, P. C., Gotham, K., Guthrie, W., Luyster, R. J., Risi, S., & Rutter, M. (2012). *Autism diagnostic observation schedule : ADOS-2*: Western Psychological Services.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., . . . 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.

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.

Mares, I., Smith, M. L., Johnson, M. H., & Senju, A. (2016). Direct gaze facilitates rapid orienting to faces: Evidence from express saccades and saccadic potentials. *Biological psychology, 121*, 84-90.

Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jeffries, H., & Plaisted, K. (2002). High motion coherence thresholds in children with autism. *Journal of Child Psychology and Psychiatry, 43*(2), 255-263.

Milton, D. E. (2012). On the ontological status of autism: the 'double empathy problem'. *Disability & Society, 27*(6), 883-887.

Mirenda, P. L., Donnellan, A. M., & Yoder, D. E. (1983). Gaze behavior: A new look at an old problem. *Journal of Autism and Developmental Disorders, 13*(4), 397-409.

Moore, C., & Barresi, J. (2017). The role of second-person information in the development of social understanding. *Frontiers in Psychology, 8*, 1667.

Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology and Psychiatry, 44*(6), 793-809.

Mundy, P. (2016). *Autism and joint attention: Development, neuroscience, and clinical fundamentals*: Guilford Publications.

Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience, 47*(6), 497-514.

Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks, 23*(8-9), 985-997.

Mundy, P., & Neal, A. R. (2000). Neural plasticity, joint attention, and a transactional social-orienting model of autism. In *International review of research in mental retardation* (Vol. 23, pp. 139-168): Elsevier.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current directions in psychological science, 16*(5), 269-274.

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.

Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. *Development and psychopathology, 6*(3), 389-401.

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 and other developmental disabilities, 23*(1), 5-14.

Myllyneva, A., & Hietanen, J. K. (2015). There is more to eye contact than meets the eye. *Cognition, 134*, 100-109.

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: is it normal? Is it automatic? Is it social? *Development and psychopathology, 20*(1), 79-97.

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

Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and psychopathology, 14*(2), 239-251.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in cognitive sciences, 16*(10), 504-510.

Pellicano, E., Gibson, L., Maybery, M., Durkin, K., & Badcock, D. R. (2005). Abnormal global processing along the dorsal visual pathway in autism: a possible mechanism for weak visuospatial coherence? *Neuropsychologia, 43*(7), 1044-1053.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology, 32*(1), 3-25.

Protopapas, A. (2014, June 2015). An introductory explanation of contrast coding in R linear models. Retrieved from https://rstudio-pubs-static.s3.amazonaws.com/84177_4604ecc1bae246c9926865db53b6cc29.html

Quené, H., & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication, 43*(1-2), 103-121.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of memory and language, 59*(4), 413-425.

Redcay, E., Dodell-Feder, D., Mavros, P. L., Kleiner, M., Pearrow, M. J., Triantafyllou, C., . . . Saxe, R. (2013). Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. *Human brain mapping, 34*(10), 2511-2523.

Redcay, E., Kleiner, M., & Saxe, R. (2012). Look at this: the neural correlates of initiating and responding to bids for joint attention. *Frontiers in Human Neuroscience, 6*, 169.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, 1.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *Cran R*.

Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. *Current directions in psychological science, 25*(1), 70-74.

Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire*. Los Angeles: Western Psychological Services.

Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., . . . Fujibayashi, Y. (2010). "Stay tuned": inter-individual neural synchronization during mutual gaze and joint attention. *Frontiers in integrative neuroscience, 4*, 127.

Sander, D., Grandjean, D., Kaiser, S., Wehrle, T., & Scherer, K. R. (2007). Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion. *European Journal of Cognitive Psychology, 19*(3), 470-480.

Sasson, N. J., Faso, D. J., Nugent, J., Lovell, S., Kennedy, D. P., & Grossman, R. B. (2017). Neurotypical peers are less willing to interact with those with autism based on thin slice judgments. *Scientific Reports, 7*, 40700.

Saxe, R. (2006). Uniquely human social cognition. *Current opinion in neurobiology, 16*(2), 235-239.

Schilbach, L. (2010). A second-person approach to other minds. *Nature Reviews Neuroscience, 11*(6), 449.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and brain sciences, 36*(4), 393-414.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., . . . Vogeley, K. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *Journal of cognitive neuroscience, 22*(12), 2702-2715.

Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia, 44*(5), 718-730.

Scott-Phillips, T. C. (2018). Cognition and communication. *The International Encyclopedia of Anthropology*, 1-10.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology, 18*(9), 668-671.

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition, 108*(2), 303-319.

Senju, A., & Johnson, M. H. (2009a). Atypical eye contact in autism: models, mechanisms and development. *Neuroscience & Biobehavioral Reviews, 33*(8), 1204-1214.

Senju, A., & Johnson, M. H. (2009b). The eye contact effect: mechanisms and development. *Trends in cognitive sciences, 13*(3), 127-134.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science, 325*(5942), 883-885.

Senju, A., Tojo, Y., Dairoku, H., & Hasegawa, T. (2004). Reflexive orienting in response to eye gaze and an arrow in children with and without autism. *Journal of Child Psychology and Psychiatry, 45*(3), 445-458.

Senju, A., Yaguchi, K., Tojo, Y., & Hasegawa, T. (2003). Eye contact does not facilitate detection in children with autism. *Cognition, 89*(1), B43-B51.

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition, 189*, 260-274.

Sodian, B., Schuwerk, T., & Kristen, S. (2015). Implicit and spontaneous theory of mind reasoning in autism spectrum disorders. *Autism spectrum disorder—Recent advances*, 113-135.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142): Harvard University Press Cambridge, MA.

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of cognitive neuroscience, 21*(3), 489-510.

Swettenham, J., Condie, S., Campbell, R., Milne, E., & Coleman, M. (2003). Does the perception of moving eyes trigger reflexive visual orienting in autism? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1430), 325-334.

Tomasello, M. (1995). Joint attention as social cognition. *Joint attention: Its origins and role in development, 103130*.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental science, 10*(1), 121-125.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and brain sciences, 16*(3), 495-511.

Ulvund, S. E., & Smith, L. (1996). The predictive validity of nonverbal communicative skills in infants with perinatal hazards. *Infant Behavior and Development, 19*(4), 441-449.

Van Boxtel, J. J., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology, 4*, 19.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological review, 121*(4), 649.

Vermeulen, P. (2015). Context blindness in autism spectrum disorder: Not using the forest to see the trees as trees. *Focus on autism and other developmental disabilities, 30*(3), 182-192.

Vivanti, G., Fanning, P., Hocking, D., Sievers, S., & Dissanayake, C. (2017). Social Attention, Joint Attention and Sustained Attention in Autism Spectrum Disorder and Williams Syndrome: Convergences and Divergences. *Journal of Autism and Developmental Disorders, 47*(6), 1866-1877. doi:10.1007/s10803-017-3106-4

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*: Harvard university press.

Wallace, S., Coleman, M., Pascalis, O., & Bailey, A. (2006). A study of impaired judgment of eye-gaze direction and related face-processing deficits in autism spectrum disorders. *Perception, 35*(12), 1651-1664.

Wilson, D., & Sperber, D. (2002). Relevance Theory. In L. R. Horn & G. Ward (Eds.), *Handbook of Pragmatics*. Oxford: Blackwell.

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.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International journal of developmental neuroscience, 23*(2-3), 143-152.

Appendix

Office of the Deputy Vice-Chancellor (Research)

Research Services
Research Hub, 17 Wally's Walk
Macquarie University
NSW 2109 Australia
T: +61 (2) 9850 7987
http://www.research.mq.edu.au/
ABN 90 952 801 237
CRICOS Provider No 00002J

13/12/2018

Dear Dr Nathan Caruana,

Reference No:5201837756676

Title: 3775 Using virtual reality to understand how autistic people interact with others

Thank you for submitting the above application for ethical and scientific review. Macquarie University Human Research Ethics Committee HREC Humanities & Social Sciences considered your application.

I am pleased to advise that ethical and scientific approval has been granted for this project to be conducted by Dr Nathan Caruana and other personnel: Professor Genevieve McArthur, Dr David Kaplan, Professor Elizabeth Pellicano, Ayeh Alhasan, Professor Michael Richardson, Dr Patrick Nalepka, Miss Christine Inkley, Ms Hannah Rapaport**.**

Approval Date: 13/12/2018

This research meets the requirements set out in the *National Statement on Ethical Conduct in Human Research* (2007, updated July 2018) (the *National Statement*).

Standard Conditions of Approval:

1. Continuing compliance with the requirements of the *National Statement,* which is available at the following website: http://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research
2. This approval is valid for five (5) years, subject to the submission of annual reports. Please submit your reports on the anniversary of the approval for this protocol.
3. All adverse events, including events which might affect the continued ethical and scientific acceptability of the project, must be reported to the HREC within 72 hours.
4. Proposed changes to the protocol and associated documents must be submitted to the Committee for approval before implementation.

It is the responsibility of the Chief investigator to retain a copy of all documentation related to this project and to forward a copy of this approval letter to all personnel listed on the project.

Should you have any queries regarding your project, please contact the Ethics Secretariat on 9850 4194 or by email ethics.secretariat@mq.edu.au

The HREC Terms of Reference and Standard Operating Procedures are available from the Research Office website at: https://www.mq.edu.au/research/ethics-integrity-and-policies/ethics/human-ethics

The HREC Humanities & Social Sciences wishes you every success in your research.

Yours sincerely,

Dr Karolyn White
Chair, HREC Humanities & Social Sciences

This HREC is constituted and operates in accordance with the National Health and *Medical Research Council's (NHMRC)National Statement on Ethical Conduct in Human Research (2007, updated July 2018*) and the *CPMP/ICH Note for Guidance on Good Clinical Practice*