Evidence for the adaptive parsing of non-communicative eye movements during joint attention interactions

Ayeh Alhasan [Corresp., 1, 2], **Nathan Caruana** [Corresp. 1, 2]

1 School of Psychological Sciences, Macquarie University, Sydney, New South Wales, Australia

2 Perception in Action Research Centre, Macquarie University, Sydney, New South Wales, Australia

Corresponding Authors: Ayeh Alhasan, Nathan Caruana
Email address: ayeh.alhasan@mq.edu.au, nathan.caruana@mq.edu.au

During natural social interaction, the ability to detect and respond to gaze-based joint attention bids often involves the evaluation of non-communicative eye movements. However, very little is known about how much humans are able to track and parse spatial information from these non-communicative eye movements over time, and the extent to which this influences joint attention outcomes. This was investigated in the current study using an interactive computer-based joint attention game. Using a fully within-subjects design, we specifically examined whether participants were quicker to respond to communicative joint attention bids that followed predictive, as opposed to random or no, non-communicative gaze behaviour. Our results suggest that in complex, dynamic tasks, informative gaze behaviour is tracked and used to plan subsequent joint attention responses. We also went further to examine the extent to which this ability to track dynamic spatial information was specific to processing gaze information. This was achieved by comparing performance to a closely matched non-social task that comprised of dynamic arrow stimuli. Whilst we found that people are also able to track and use dynamic non-social information from arrows, there was clear evidence for a relative advantage for tracking gaze cues during social interactions. The implications of these findings for social neuroscience and autism research are discussed.

Evidence for the Adaptive Parsing of Non-Communicative Eye Movements during Joint Attention Interactions

Ayeh Alhasan[1][2], Nathan Caruana[1][2]

[1] School of Psychological Sciences, Macquarie University, Sydney, New South Wales, Australia.
[2] Perception in Action Research Centre, Macquarie University, Sydney, New South Wales, Australia

Corresponding Authors:
Ayeh Alhasan, Nathan Caruana

School of Psychological Sciences, Faculty of Medicine, Health and Human Sciences, Macquarie University, 16 University Ave., Sydney, New South Wales 2109, Australia.
Email address: ayeh.alhasan@mq.edu.au, nathan.caruana@mq.edu.au

Abstract

During social interactions, the ability to detect and respond to gaze-based joint attention bids often involves the evaluation of non-communicative eye movements. However, very little is known about how much humans are able to track and parse spatial information from these non-communicative eye movements over time, and the extent to which this influences joint attention outcomes. This was investigated in the current study using an interactive computer-based joint attention game. Using a fully within-subjects design, we specifically examined whether participants were quicker to respond to communicative joint attention bids that followed predictive, as opposed to random or no, non-communicative gaze behaviour. Our results suggest that in complex, dynamic tasks, people adaptively *use* and *dismiss* non-communicative gaze information depending on whether it informs the locus of an upcoming joint attention bid. We also went further to examine the extent to which this ability to track dynamic spatial information was specific to processing gaze information. This was achieved by comparing performance to a closely matched non-social task where eye gaze cues were replaced with dynamic arrow stimuli. Whilst we found that people are also able to track and use dynamic non-social information from arrows, there was clear evidence for a relative advantage for tracking gaze cues during social interactions. The implications of these findings for social neuroscience and autism research are discussed.

Introduction

Joint attention is the 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). The early development of joint attention is pivotal in supporting the later development of language (Akhtar et al., 1991; Charman, 2003; Dawson et al., 2004) and social-cognitive skills, including the ability to understand and predict the mental states and perspectives of others (Mundy, 2003, 2016, 2018; Mundy & Jarrold, 2010; Mundy & Neal, 2000; Mundy & Newell, 2007).

Until recently, most experimental studies of 'joint attention' required participants to observe and respond to a single gaze cue on each trial, often in a non-interactive context (see Caruana et al., 2017b; Frischen et al., 2007; Nation & Penny, 2008 for reviews). However, in genuine daily interactions gaze cues are observed in the context of a *dynamic* and experiential flow of perception and action (Heft, 2003). The dynamic nature of gaze requires us to parse constant streams of eye movements to determine when a gaze shift is communicative and intentionally signals an opportunity for joint attention. According to the Relevance Theory and the 'ostensive-inferential communication' model, this differentiation is achieved by identifying the relevance of ostensive signals (e.g., eye contact) and using them to infer the communicative intent of a social partner (Sperber & Wilson, 1986; Wilson & Sperber, 2002). This notion of relevance is not limited to the transmission of communicative intentions. Rather, it is a more general property of information processing that can be applied to other categories of information, including non-communicative but informative gestures. In the context of gaze-based joint attention, it is not yet clear whether and to what extent people evaluate the relevance of non-communicative eye movements in an interactive context. As such, the current study aimed to examine whether, and to what extent, humans parse and are influenced by the non-communicative eye movements that precede eye contact and subsequent bids for joint attention. This endeavour is critical for fully understanding how gaze signals are used to guide social coordination in realistic dynamic interactions.

Ostensive Eye Contact

Ostensive signals such as eye contact are argued to signal communicative intent, which can increase the perceived relevance and communicativeness of subsequent behaviours (including eye movements; Sperber & Wilson, 1986; Wilson & Sperber, 2002). The ability of eye contact to capture attention is observed in infants within their first year, and is believed to play a crucial role in signaling relevant social information during interactions (Behne, Carpenter, Tomasello, 2005; Csibra, 2010; Wilson & Sperber, 2002). Indeed, studies show that eye contact, especially during live interactions, induces psychophysiological arousal that results in increased alertness and attention (Gale et al., 1975; Helminen et al., 2011; Hietanen et al., 2020; Kleinke & Pohlen, 1971; Nichols & Champness, 1971). Other studies and related models (e.g., Fast-track modulator model; Senju & Johnson, 2009), also indicate that observing eye contact results in a rapid and

80   automatic activation of the social brain network and subsequently upregulates social-cognitive
81   processing – including the representation of others' mental states – via a subcortical pathway
82   (Conty et al., 2007; Mares et al., 2016; Senju & Johnson, 2009). However, it is not yet clear if
83   and how these eye contact signals are used to identify communicative joint attention bids when
84   embedded in a dynamic sequence of eye movements.
85
86   **Non-communicative eye movements**
87   In very recent work, we conducted the first experimental interrogation of the extent to which
88   humans process and are influenced by non-communicative eye movements leading up to a joint
89   attention opportunity (Caruana et al., 2020; Caruana et al., 2017a). This revealed that humans do
90   sensitively parse and differentiate communicative and non-communicative eye movements
91   during dynamic gaze-based interactions. Specifically, in our most recent study (Caruana et al.,
92   2020), we designed two experiments that manipulated the presence and predictability of non-
93   communicative eye gaze behaviour in the lead-up to a joint attention episode. In both
94   experiments, participants played a co-operative game where they needed to catch a burglar using
95   nothing but their gaze (adapted from Caruana et al., 2015; also see Caruana et al., 2017b for an
96   in-depth discussion and review of this methodological approach). The game was played with an
97   avatar controlled by a gaze-contingent algorithm, and participants were asked to collaborate with
98   their partner by initiating and responding to joint attention bids (see *Fig. 1* for stimuli
99   presentation). To do this, they both needed to search through their allocated row of houses to find
100 the burglar. We refer to this portion of the trial as the 'search-phase'. Critically the task was
101 programmed such that the avatar always completed their search for the burglar last. As such,
102 during the search phase participants observed their partner make a series of non-communicative
103 eye movements as they completed their search. At the end of the search phase, participants were
104 required to initiate joint attention if they found the burglar, or respond to their partner's joint
105 attention bid if they did not. Participants were not provided with any explicit instructions as to
106 how they should perform the task, but for them to do it correctly they needed to establish eye
107 contact before guiding or being guided by the avatar. We manipulated the presence and
108 predictability of the avatar's non-communicative searching eye movements across two
109 experiments and measured the time it took participants to respond to a guiding cue by making an
110 accurate saccade to the correct house.
111
112 In the first experiment, the presence of irrelevant non-communicative gaze cues was
113 manipulated. Here, we compared joint attention responsivity when the avatar displayed an
114 irrelevant non-communicative (i.e., Random Search) sequence of eye movements before
115 establishing eye contact with a simpler version of the task where the search phase was removed
116 (i.e., NoSearch). In the second experiment, the informative nature of the non-communicative
117 searching gaze pattern was manipulated to investigate its effect on joint attention responsivity.
118 Here, we compared the same Random Search condition (described above) with a more closely-
119 matched 'Predictive' search condition, in which the avatar's non-communicative search gaze was

120 predictive of the target location (i.e., Predictive Search). Specifically, the avatar was
121 programmed to look at the burglar's location last before establishing eye contact and guiding the
122 participant to that same location. This differs to the Random Search condition where the final
123 gaze shift during the avatar's search was randomly determined. Participants also completed non-
124 social versions of these tasks in which gaze cues were replaced with arrow cues. We found that
125 people were faster to respond to joint attention bids when the non-communicative eye
126 movements offered relevant (Predictive), rather than irrelevant (Random), spatial information
127 about the locus of joint attention. We also found tentative evidence that humans may be able to
128 discount irrelevant (Random) non-communicative gaze information more readily than they are
129 for non-social spatial cues such as arrows in dynamic contexts (Caruana et al., 2020).

131 The examination of dynamic gaze contexts by Caruana et al. (2020) involved two distinct
132 experiments with different participant samples. As such, from these data we are unable to
133 directly compare performance between the NoSearch and Predictive Search contexts. This is
134 critical in elucidating whether an informative non-communicative gaze context such as this leads
135 to faster response times than no gaze context at all. Another limitation of our previous work was
136 that the perceptual difference in the transition between the 'search' and 'response' phases of each
137 trial were not completely matched across social and non-social conditions. Specifically, the green
138 fixation point and green arrow stimulus in the non-social 'Arrow' condition did not match the
139 difference in perceptual salience between averted and direct gaze stimuli in the social 'Eyes'
140 condition. On Arrow trials, the green fixation point was followed by a green arrow that extended
141 from the same green fixation point. This resulted in a relatively more subtle transition between
142 'searching' and 'guiding' arrow shifts since they were intervened by a fixation point of the same
143 colour. This contrasts with the arguably more perceptually obvious transition between
144 'searching' and 'guiding' gaze shifts, which were intervened by direct gaze (i.e., eye contact). As
145 such, it is possible that the previously-reported 'relative advantage' for eye gaze was not driven
146 by a unique feature of eyes, but was rather driven by the more salient visual contrast between
147 direct and averted gaze. To address this, we adapted the control stimulus in the current work,
148 replacing the green fixation point stimulus with a yellow fixation point (see *Fig. 2B*).

150 **Current study**
151 The current study endeavoured to rigorously examine how humans parse and are influenced by
152 the non-communicative eye movements that precede eye contact and subsequent bids for joint
153 attention. Building on recent work, we compared the three contexts (Predictive, No-Search,
154 Random) using both stimulus sets (Eyes, Arrows) with a fully within-subjects design.
155 Furthermore, to ensure the reported effects were indeed unique to the social interaction, we
156 adapted the control stimulus originally implemented by Caruana et al. (2020) to better match the
157 perceptual changes observed during social trials. Based on our previous findings (Caruana et al.,
158 2020), and consistent with the ostensive-inferential communication model (Sperber & Wilson,
159 1986; Wilson & Sperber, 2002), we hypothesised that even after implementing a more closely-

160 matched control stimulus, ostensive signals (i.e., eye contact) would result in a uniquely social
161 advantage by assisting participants to selectively and rapidly respond to communicative gaze
162 shifts. This would be characterised by participants showing slower responses on the Random
163 Arrows condition but not the Random Eyes when compared to the NoSearch condition.
164 Additionally, we expect participants to show faster responses on the Predictive Search condition
165 when compared to NoSearch, with larger effects observed for Eyes than Arrows.
166

Materials & Methods

167

Ethics Statement

168
169 All procedures implemented in this study were carried out in accordance with the protocol
170 approved by the Macquarie University Human Research Ethics Committee (ID: **3775**).
171 Participants provided written, informed consent to take part in this study prior to participation.
172

Participants

173
174 Thirty-one adult participants were recruited from a pool of undergraduate psychology students at
175 Macquarie University and were given course credit for their time. All participants reported
176 normal or corrected-to-normal vision (with clear contact lenses only), and reported no history of
177 neurological injury or impairment. Two participants stated that they did not believe the deceptive
178 cover story used in this study (see below) and were excluded from all analyses. The final sample
179 included 29 participants (M_{age}= 19.76 years; SD= 4.4; 8 males).
180

Stimulus and Apparatus

181
182 Participants were seated at a table with a chin and forehead rest installed to stabilise head
183 movements and standardise the screen viewing distance. Participants played a cooperative game
184 with an on-screen avatar. The experimental stimuli were presented using Experiment Builder
185 1.10.165 on a 27-inch AOC monitor (display size: 59.8 cm x 33.6 cm; resolution: 1920 x 1080
186 pixels; refresh rate: 144 Hz) positioned 80 cm away from the participant. A remote desktop-
187 mounted Eyelink 1000 (SR Research Ltd., Ontario, Canada) was used to record eye-movements
188 from the right eye at a sampling rate of 500Hz. Before starting each block, a 9-point eye-tracking
189 calibration and validation was implemented.
190

191 During the task trials, an avatar, presented as an anthropomorphic animated face subtending
192 6.08° x 3.65° of visual angle, was displayed in the centre of the screen surrounded by six houses,
193 each subtending 3.58° of visual angle (see *Fig. 1*). There were seven invisible gaze-related areas
194 of interest (AOI) defined around the avatar's face and each of the six houses that were used by
195 our gaze-contingent algorithm during the experiment and for subsequent data processing (see
196 Caruana et al., 2015, for detailed description of AOI definition).
197

Design and procedure

198
199

200 Participants were told that they would be playing a collaborative game with two different
201 members of the research team, named 'Alan' and 'Tony', who would be interacting with them
202 from an adjacent laboratory. They were told that their partner's eyes would be recorded using an
203 eye-tracker and used to control the eye movements of the avatar they saw on their screen.
204 Participants were also told that they would control an avatar presented on Alan/Tony's screen in
205 the same way. In reality, the avatar's eye movements were controlled by a gaze-contingent
206 algorithm which responded appropriately to the participant's own eye movement behaviour. This
207 simulated the experience of a convincing and reciprocal social interaction.
208
209 Each trial began with the participant searching for the burglar in the houses with blue doors
210 located at the top of the screen, while the avatar searched houses with red doors at the bottom of
211 the screen. To search their allocated houses, participants had to look at a house before the door
212 opened to either reveal the burglar or an empty house. To help introduce variability in the spatial
213 sequence of participant's search behaviour, some trials started with one or two houses already
214 open and empty. The sequence and number of open houses were systematically varied and
215 counterbalanced across trials. This ensured that the avatar's search behaviour was realistically
216 variable and that participants could not predict which or how many doors their partner would
217 search before initiating joint attention.
218
219 On responding trials, participants would not find the burglar in any of their houses and had to
220 wait for their partner to complete his search and guide them to the correct location. Once the
221 participant completed their search and fixated on the avatar's face, it searched 1–2 more houses
222 before establishing eye contact. This ensured that participants observed the avatar's non-
223 communicative (i.e., searching) and communicative (i.e., guiding) eye gaze behaviours. This also
224 allowed us to manipulate the presence and predictability of non-communicative gaze cues
225 displayed during this searching phase. On initiating trials, participants found the burglar in one of
226 their houses and had to capture their partner's attention by establishing eye contact before
227 guiding him to the correct location. Participants were told that for them to successfully catch the
228 burglar they needed to both look at the target location. Participants were not given explicit
229 instructions as to how they should guide or be guided by their partner. Once joint attention was
230 achieved, the burglar appeared behind bars with a police car 'arriving' at the correct location to
231 provide positive feedback.
232
233 In this study, we manipulated both the spatial context preceding the joint attention bid as well as
234 the cue stimulus used to control for non-social attention effects, yielding six conditions: three
235 *context* conditions (NoSearch, Random Search and Predictive Search) and two *stimulus*
236 conditions (Eyes and Arrows). This resulted in each participant completing six condition blocks,
237 with 30 responding trials and 30 initiating trials per block. Although the primary planned
238 analyses for this study were for responding trials, it was essential to retain initiating trials in the
239 task as they were critical for establishing the reciprocal joint attention task context.

240

241 The primary conditions were arranged into blocks, with block/condition order counterbalanced
242 across participants. However, for each participant, the two stimulus conditions within each
243 context condition were administered consecutively, in counterbalanced order, to minimise the
244 switching between spatial cue contexts (e.g., Eyes Random Search, Arrows Random Search;
245 Eyes Predictive Search, Arrows Predictive Search; Eyes NoSearch, Arrows NoSearch). Within
246 each condition block, trial order was randomised to ensure that the location of the burglar and the
247 number of gaze shifts made by the avatar were not conflated with any potential order effects.
248 Participants could make four types of errors on each trial: (1) Search errors occurred when
249 participants spent more than 3000 ms looking away from the avatar or their houses during the
250 search-phase, which resulted in the text "Failed Search" appearing on the screen; (2) Timeout
251 error occurred when participants took over 3000 ms to respond to the relevant gaze-cue; and (3)
252 Location errors occurred when participants responded by looking at an incorrect location (see
253 Results for detailed breakdown of error data). For both Timeout and Location errors, the burglar
254 appeared in red at the target location to provide negative feedback. Error trials were excluded
255 from subsequent analyses of Saccadic Reaction Time (SRT) data.

256

257 **Context Conditions**
258 We manipulated the gaze information displayed during the non-communicative search-phase of
259 two context conditions, (Random and Predictive) and a third 'NoSearch' baseline context
260 condition that did not comprise any non-communicative information. See *Fig. 2A* for a summary
261 of the trial sequence for each of these conditions.
262   **Random Search condition.** In this condition, when participants completed their search
263 and were waiting for 'Alan' to establish eye contact and initiate joint attention, the final gaze
264 shift made by Alan before establishing eye contact was randomly determined. This is consistent
265 with the original implementations of this task (Caruana et al., 2015; Caruana et al., 2018) which
266 ensured participants could not predict where Alan was going to guide them based on the final
267 target searched.
268   **Predictive Search condition.** This condition was identical to the Random Search
269 condition, with the exception that the final house searched by the avatar before establishing eye
270 contact was always the target location, thus predicting the upcoming locus of joint attention. To
271 prevent participants from detecting this systematic difference in avatar behavior across the
272 Predictive and Random contexts, a cover story was implemented stating that participants would
273 be playing the game with another member of the research team named 'Tony.' It is worth noting
274 that participants were not explicitly informed of this manipulation. To verify whether
275 participants detected any systematic difference between the Random and Predictive context
276 conditions, we performed a post-experimental interview to collect descriptive information about
277 their subjective experiences. Indeed, 6 people reported noticing the difference between Alan and
278 Tony's search patterns (For a full summary of the subjective data, see documentation on our OSF
279 Project page https://osf.io/e7kg8/).

280   **NoSearch condition.** In this condition, participants completed a simpler version of the
281   task with Alan in which there was no need to search for the burglar on each trial. On initiating
282   trials, participants could see only one blue door remaining closed, which they knew to conceal
283   the burglar. On these trials, they simply waited for Alan to open his eyes to establish eye contact
284   and initiated joint attention using a single gaze shift. Similarly on responding trials, Alan opened
285   his eyes to establish eye contact and guided participants to the correct location using a single
286   gaze shift.
287
288   **Stimulus Conditions**
289   A matched control condition was completed as separate blocks for each context condition. This
290   was implemented to control for non-social task demands, such as attentional, oculomotor and
291   inhibitory control. This resulted in two stimulus conditions (Eyes, Arrows) using the same gaze-
292   contingent algorithm for presenting stimuli in both conditions. As such the spatial and temporal
293   dynamics in these two conditions were identical. For the arrow condition, participants were
294   informed that they were completing a computer-simulated version of the task in which a
295   computer-controlled arrow stimulus was used to guide them to the correct location. The avatar's
296   face remained on the screen with its eyes closed to best match the visual context across all
297   conditions. At the beginning of each block, a grey fixation point subtending a visual angle of
298   0.29° was presented in between the avatar's eyes (analogous to the avatar closing its eyes on
299   social gaze trials). The fixation point then turned yellow (analogous to the avatar displaying
300   direct eye gaze). This was followed by a green arrow extending from a green central fixation
301   point subtending a visual angle of 1.08°. Critically, this was analogous of the avatar averting its
302   gaze. During the search phase of the Random and Predictive Search conditions, the arrow
303   stimulus was updated to point at different houses to match the avatar's searching behaviour
304   within the social condition. Once participants completed their search and looked back at the
305   central area of interest (AOI), the arrow stimulus pointed at 1-2 more houses before being
306   replaced by the yellow fixation point, analogous of eye contact. This was then followed by a
307   single green arrow pointing towards the target house which participants needed to follow to
308   successfully 'catch the burglar'.
309
310   **Statistical Analyses**
311   Interest area and trial reports were exported using DataViewer software (SR Research Ltd.,
312   Ontario, Canada) to analyse the accuracy and Saccadic Reaction Time (SRT) data. For accuracy
313   analyses, 'Calibration' and 'Search' errors were removed before analysing the remaining trials
314   for the proportion of correct trials. This was done because these errors occurred before the
315   relevant gaze or arrow cue was presented and, hence, do not represent true joint attention errors.
316   For SRT analyses we only included correct trials. Trials in which participants responded faster
317   than 150 ms were also excluded as these were likely to be anticipatory responses (Carpenter,
318   1988). Raw eye-tracking data was screened and analysed in R using a custom script
319   (https://osf.io/e7kg8/).

320
321   Consistent with our previous studies (Caruana et al., 2020; Caruana et al., 2017a) logistic and
322   linear mixed random effects (LME) were used to analyse accuracy and SRT respectively.
323   Specifically, we wanted to evaluate evidence for effects of Context, Stimulus and their
324   interaction. The maximum likelihood estimation method was implemented in these analyses
325   using the lme4 R package (Bates & Sarkar, 2005) and p-values were estimated using the
326   lmerTest package (Kuznetsova et al., 2015).
327
328   We were interested in investigating the main effect of stimulus (Eyes, Arrows), context
329   (Random, NoSearch, Predictive), and their interaction. To do this we used the 'successive
330   differences contrast coding' method within the 'MASS' package in R (Ripley et al., 2013). This
331   contrast method estimates effect parameters by sequentially comparing each level of context with
332   the next level specified in the model. This method was used to provide parameter estimates for
333   the overall effect of stimulus as well as the context effects between (1) NoSearch and Random
334   Search, and (2) NoSearch and Predictive Search. Of critical interest in the current study, we were
335   interested in testing the interaction of context and stimulus effects. As such, parameter estimates
336   were also obtained for the context-by-stimulus interaction in (1) and (2) above.
337
338   We also tested the effect of context between the Predictive and Random Search conditions as
339   well as their interaction with stimulus. However, this contrast could not be estimated using the
340   predefined successive differences contrast coding method. This is because in this coding method
341   the contrast coefficients are chosen so that the coded coefficients are the differences between the
342   means of the second and first levels, the third and second levels, and so on. Therefore, we used
343   the 'emmeans' package to manually define the missing comparisons (Lenth et al., 2019). Finally,
344   we ran post-hoc analyses for accuracy and SRT data to assist in interpreting significant stimulus-
345   by-context interaction effects. An FDR correction was then applied to these post-hoc contrasts to
346   confirm significance after correcting for multiple comparisons (Benjamini & Hochberg, 1995).
347   Accuracy and SRT models were defined with maximally-defined random-factor structures,
348   including random intercepts for trial, block order and by-subject random slopes for the intercept
349   and fixed effects (Barr et al., 2013). For SRT analyses, the residuals of the raw data violated the
350   normality assumption and hence data were transformed using an inverse transformation. The
351   normality assumption was confirmed after applying the transformation (details can be found in
352   accompanying R code and output (https://osf.io/e7kg8/); see Balota et al., 2013). All analyses
353   had a significance criterion of $\alpha = 0.05$.
354
355   For estimating effect-size, the chi-squared goodness-of-fit tests were performed comparing a
356   number of mixed random-effects models using Chi-square likelihood ratios to quantify the
357   contribution of each fixed effect and interaction parameter to the model fit (Johnston et al.,
358   2006). Unlike traditional measures of effect-size (e.g., r2), this method provides an estimation of
359   the variance explained by each fixed effect while accounting for variance independently

360 explained by the specified random effects (see Caruana et al., 2021; Caruana et al., 2019). For
361 full details pertaining to all of the analyses described above, refer to the accompanying R code
362 (https://osf.io/e7kg8/).
363
Results

365

Accuracy.

367 Overall, only 9.30% of trials involved error, with the majority being Location errors (5.29% of
368 trials), followed by Search (2.45%) and Timeout errors (1.55%). Descriptive statistics of
369 estimated fixed effect parameters are summarised in *Table 1*. Accuracy data for context and
370 stimulus conditions are illustrated in *Fig. 3*.
371
372 Overall, participants made significantly more errors on the Random than NoSearch trials (β =
373 1.90, z = 5.34, p <.001). However, we found no evidence for an effect of stimulus (β = -0.06, z =
374 -0.19, p = 0.85) or context between the Predictive and NoSearch contexts (β = -0.66, z = -1.43, p
375 = 0.15). There was also no stimulus-by-context interactions between NoSearch and either the
376 Random (β = -0.83, z = -1.83, p = 0.07) or Predictive (β = -0.07, z = -0.13, p = 0.89) conditions.
377 We used emmeans to compare the accuracy outcomes in the Random and Predictive Search
378 contexts (see Statistical Analysis above for justification). This revealed a significant context
379 effect (β = -2.40, z = -3.74, p <.001) and stimulus-by-context interaction (β = -0.89, z = -2.34, p
380 = 0.02). Post-hoc pairwise comparisons investigating this stimulus-by-context interaction
381 between Random and Predictive contexts were then defined, correcting for multiple comparisons
382 using a false discovery rate (FDR) correction. This revealed a significant effect of context
383 between Random and Predictive with a larger effect for Arrows (β = -1.65, z = -4.31, p <.000)
384 than Eyes (β = -0.75, z = -2.06, p = 0.04).
385

Saccadic Reaction Times.

387 Of critical importance in the current study, we investigated whether and how the presence of
388 random and predictive spatial signals differentially influenced the speed with which participants
389 were able to prepare saccadic responses to eye gaze and arrow cues. Mean SRT are summarised
390 by condition in *Table 2* and descriptive statistics of estimated fixed effect parameters are
391 summarised in *Table 3*. All SRT data are summarised by context and stimulus conditions in *Fig.*
392 *4*.
393
394 Participants were significantly faster to respond during Predictive than NoSearch trials (β = 0.64,
395 t = 5.87, p < .000). However, we found no evidence for a significant difference between the
396 Random and NoSearch contexts (β = 0.15, t = 1.98, p = 0.06). We also found that, overall,
397 participants were faster to respond to Arrows than Eyess (β = -0.16, t = -2.02, p = 0.05).
398 Additionally, we found significant stimulus-by-context interactions when comparing each of the

399   Random and Predictive contexts to the NoSearch context (Random: β = -0.27, t = -4.08, p <
400   .000; Predictive: β = 0.20, t = 2.91, p < 0.01). The context effect and the stimulus-by-context
401   interaction between Random and Predictive contexts were also estimated using emmeans and
402   revealed significantly slower responses on the Random context than the Predictive context (β = -
403   1.57, t = -8.78, p < .000) but no significant interaction (β = -0.07, t = -1.00, p = 0.32). We used
404   emmeans to conduct post-hoc pairwise comparisons of interest using custom contrasts, and
405   correcting for multiple comparisons using a FDR correction, to help understand the significant
406   interaction effects. This analysis revealed a significant effect of stimulus in the NoSearch context
407   (β = 0.31, t = 3.71, p = .001) but not the Random (β = 0.04, t = 0.46, p = 0.75) or Predictive
408   contexts (β = 0.11, t = 1.27, p = 0.29). We also found a significant effect of context between
409   Random and NoSearch with Arrows (β = -0.29, t = -3.45, p = .002) but not Eyes (β = -0.01, t = -
410   0.16, p = 0.88). Further, there was a significant effect of context between Predictive and
411   NoSearch with a larger effect for Eyes (β = -0.73, t = -6.48, p < .000) than Arrows (β = -0.53, t =
412   -4.70, p < .000).
413
414   **Model fit analyses.** For quantifying the effects of stimulus and context, model-fit-improvement
415   was compared as a function of each fixed effect parameter. Compared to the null model (i.e., a
416   model with no fixed-effect factors), adding the context factor significantly improved the model
417   fit by 36.05 times ($X^2(1)$ = 36.05, p < 0.000). Adding the stimulus factor to the context-only
418   model improved the model fit another 5.50 times ($X^2(1)$ = 5.50, p = 0.019). On the other hand,
419   including the stimulus factor to the null model first enhanced the model's fit by only 5.71 times
420   ($X^2(1)$ = 5.71, p = .017), while adding the context effect to the stimulus-only model significantly
421   improved the model fit 35.84 times ($X^2(1)$ = 35.84, p < .000). Critically, compared to a model
422   containing fixed-effect factors for both stimulus and context, adding the interaction parameter
423   significantly improved the model fit by 20.09 times ($X^2(1)$ = 20.09, p < .000). These analyses
424   show a markedly larger effect of context than stimulus. However, it also suggests that both
425   factors explain unique variance in the data and that the data are significantly better explained by
426   a model that specifies a stimulus-by-context interaction.
427

Discussion

429   Using a fully within-subjects design, we tested whether joint attention response times differ
430   when following communicative joint attention bids that are preceded by: (1) ***predictive*** (i.e.,
431   informative) non-communicative eye movements; (2) ***random*** (i.e., uninformative) non-
432   communicative eye movements; or (3) ***no eye movements at all***. We further examined whether
433   any observed differences are specific to the social domain by comparing performance in an
434   analogous task in which eye movement cues were replaced with dynamic arrow stimuli. We
435   found that participants did effectively track and use predictive information embedded in other's
436   eye movements before a joint attention bid; characterized by faster joint attention responses
437   following ***predictive*** rather than ***no*** non-communicative eye movement sequences. Whilst similar
438   patterns were observed for arrow stimuli, we present evidence for a relative advantage for

439 utilizing predictive information embedded in eye movements. This was further supported by the
440 finding of a larger cost of observing *random* spatial information embedded in dynamic **arrow**
441 sequences, but not for **eyes**. This demonstrates that participants were better able to disregard
442 uninformative spatial information when it was conveyed as non-communicative eyes than
443 arrows.
444

445 **Context Effects (Random vs Predictive vs NoSearch)**
446 In this study, we found that participants were faster to respond within the Predictive context than
447 both the NoSearch and Random contexts, irrespective of the task stimuli used (Eyes or Arrows).
448 That is, we observed slower reaction times in the Random than Predictive condition, even though
449 the contexts were matched for their complexity and dynamics. We also observed slower reaction
450 times for the NoSearch than Predictive condition, even though the Predictive context was
451 arguably more complex, with additional demands on attention. Together, these findings highlight
452 our capacity to extract and use predictive social and non-social spatial cues to optimise attention
453 allocation. This aligns with claims within the Relevance Theory, that cognition is geared towards
454 allocating attention to information that is maximally relevant in a given context (Sperber &
455 Wilson, 1986). The notion of maximal relevance refers to attending to and processing
456 information that produces the greatest cognitive effect for the least effort when compared to any
457 other inputs available at the time. In the context of our task, when relevant information about the
458 location of the burglar was available, participants were sensitive to it and used it to enhance their
459 responsivity, compared to contexts with only irrelevant information or no information at all.
460

461 Interestingly, reaction times in the Random context did not significantly differ from the
462 NoSearch context. This contrasts with our earlier findings, where we reported slower responses
463 on the Random context compared to the NoSearch context (Caruana et al., 2020). This non-
464 replicated effect may be attributed to the implementation of a more closely-matched non-social
465 stimuli, which in the current study made it easier to perceptually distinguish between irrelevant
466 and relevant non-social arrow cues in the Random condition (discussed further below).
467

468 **Context by Stimulus Interactions.** Of critical importance in the current study, we found that
469 context effects were modulated by stimulus type. Specifically, when comparing Random and
470 NoSearch conditions, we observed a specific disadvantage for Random trials when responding to
471 Arrows but not Eyes. Further, when comparing Predictive and NoSearch trials, we observed a
472 relatively larger advantage for Eyes than Arrows. Together, these findings suggest that people
473 are better positioned to *adaptively use* non-communicative gaze information when it offers
474 opportunities to facilitate joint attention, and *disregard* this information when it does not, than
475 they are for non-social spatial cues in analogous task contexts.
476

477 There are two possible explanations for this gaze advantage. First, it is possible that our rich
478 experiences in gaze-based communication and social interaction have led to the development of

479 expertise in evaluating the relevance of eye movements and rapidly discerning whether they are
480 communicative, informative or irrelevant. As suggested by Caruana et al., (2020), this expertise
481 in evaluating the relevance of non-communicative social behaviour from gaze might not readily
482 generalise to non-social stimuli. Indeed, in the current study, participants may have exhibited a
483 relative disadvantage for responding to arrows in the Random context because they have less
484 experience disregarding irrelevant arrow cues. They may have also exhibited a smaller relative
485 advantage for responding to Arrows in the Predictive context than Eyes due to the lack of
486 experience in parsing dynamic arrow sequences. This is unsurprising since arrows are rarely, if
487 ever, irrelevant or presented in dynamic contexts. By contrast, humans have continuous exposure
488 to social interactions involving a steady stream of mostly non-communicative eye movements
489 displayed by interlocutors.
490
491 A second explanation for this gaze advantage might be that humans have an innate endowment
492 to evaluate and associate ostension to non-verbal social gestures such as eye contact (Sperber &
493 Wilson, 1986; Wilson & Sperber, 2002). Social interaction requires the ability to distinguish
494 between communicative and non-communicative gestures, which involves identifying
495 communicative intent. This was afforded in the social condition given that non-communicative
496 and communicative eye movements were intervened by eye-contact; an ostensive signal for
497 communicative intent. By contrast, the arrows presented in the non-social control task required
498 participants to differentiate between relevant and irrelevant cues, but did not include the
499 identification of communicative intent. The presence of a relative social advantage for Eyes over
500 Arrows suggests that identifying communicative intent from eye-contact facilitated more
501 efficient joint attention responsivity. This might be, in part, because eye contact is an inherently
502 potent social signal that rapidly activates cortical mechanisms involved in making inferences
503 about others' intentions (Burra et al., 2019; Conty et al., 2007; Kampe et al., 2003; Mares et al.,
504 2016; Schilbach et al., 2006; Senju & Johnson, 2009). Indeed, eye contact, especially during live
505 interactions, induces psychophysiological arousal and increased alertness and attention (Gale et
506 al., 1975; Helminen et al., 2011; Hietanen et al., 2020; Kleinke & Pohlen, 1971; Nichols &
507 Champness, 1971). This explanation also aligns with past work using gaze cueing paradigms
508 which have shown greater gaze cueing effect after the observation of eye contact (Xu, Zhang, &
509 Geng, 2018; Kompatsiari, Ciardo, Tikhanoff, Metta, & Wykowska, 2018). More research is
510 needed to further probe the features of eye contact that can further facilitate the perception of
511 communicative intent during dynamic interactions, by manipulating variables such as the
512 presence, frequency and duration of eye contact in the lead up to joint attention opportunities.
513 The knowledge gained from such studies will inform how we can engineer optimally
514 communicative signals in the design of artificial agents for human social interaction (e.g., robots
515 and virtual characters).
516
517 **Stimulus Effect (Eyes vs Arrows)**

518   When examining the overall effects of stimulus, we did not find a significant difference in
519   reaction times between Eyes and Arrows within the Random or Predictive contexts. This differs
520   from our earlier findings in which we observed faster responses to Arrows compared to Eyes
521   within the Predictive context (Caruana et al., 2020). As previously mentioned, one key difference
522   between the current study and Caruana et al. (2020) is the modification of our non-social stimuli
523   to better match the perceptual change in stimuli that occurs when the avatar established eye
524   contact (i.e., averted gaze [search] → **eye contact** → averted gaze [cue]). The non-social arrow
525   and fixation point stimulus used in our previous studies presented an arguably less obvious
526   transition between the search and response phase (i.e., green fixation point + green arrow
527   [search] → **green fixation point** → green fixation point + green arrow [cue]) compared to the
528   ostensive eye contact stimulus in the social condition. As such, it is possible that we previously
529   saw slower reaction times to eye gaze in the Predictive context because the eye contact stimulus
530   was more visually salient than the fixation point. Thus, participants may have taken longer to
531   disengage and respond to the subsequent joint attention bid. The current study, however,
532   modified the non-social stimuli by changing the colour of the fixation point that was analogous
533   to eye contact, to provide a better visual match across the Eye and Arrow conditions. Our current
534   results indicate that this did indeed provide a better match for the perceptual salience of stimuli,
535   given that we observed more commensurate response time for both Predictive Eyes and Arrows.
536
537   It is noteworthy, however, that despite implementing more perceptually matched stimuli in the
538   current study, we still find evidence for significantly slower reaction times to Eyes than Arrows
539   in the NoSearch task, consistent with our previous findings. As previously discussed, eye contact
540   has a cascading effect on downstream attention and social cognition processing (Gale et al.,
541   1975; Helminen et al., 2011; Hietanen et al., 2020; Kleinke & Pohlen, 1971; Nichols &
542   Champness, 1971). This may increase cognitive processing load by automatically engaging
543   higher-order social-cognitive processing (Burra et al., 2019; Conty et al., 2007; Kampe et al.,
544   2003; Mares et al., 2016; Schilbach et al., 2006; Senju & Johnson, 2009). As such, the presence
545   of direct gaze during non-dynamic tasks may contribute to slower responses to Eyes than Arrows
546   due to slower disengagement from direct gaze stimuli. This is uniquely seen in the NoSearch
547   context where the trial begins with eye contact – adding to its salience – and where eye contact
548   does not serve to differentiate communicative from non-communicative gaze. As demonstrated
549   in the current study, characterizing this fundamental difference in responding to stand-alone gaze
550   and arrow stimuli in non-dynamic contexts (i.e., the NoSearch condition) offers a useful
551   'baseline' that can help interpret differences in stimuli, or the lack of, when interpreting
552   additional differences that may emerge in more dynamic contexts (e.g., Random and Predictive
553   conditions). The current findings also demonstrate the potential inadequacies of previous gaze
554   following paradigms of gaze processing and joint attention that have presented participants with
555   a single gaze cue on each trial (see Caruana et al., 2017b for review), given that the NoSearch
556   context – which most closely resembles this common approach - produced very different

557 stimulus effects to those observed in the more dynamic, and ecologically-valid contexts (e.g.,
558 Random, Predictive).

559

560 **Implications and Future Application.** The current findings, and our updated paradigm, offers a
561 new tool for carefully examining social information processing in dynamic contexts. This
562 approach also offers opportunities for studying joint attention difficulties experienced by autistic
563 people (e.g. Charman, 2003; Hobson & Hobson, 2007; Mundy et al., 1994), who often report
564 marked subjective difficulty in establishing eye contact with others (Trevisan et al., 2018; also
565 Adrien et al., 1993; Kanner, 1943; Mirenda et al., 1983; Zwaigenbaum et al., 2005). Specifically,
566 our paradigm can be used to interrogate why – and under what conditions – autistic individuals
567 may experience difficulties responding to gaze-cued joint attention bids. Early studies have
568 suggested this may be due to a difficulty in differentiating communicative and non-
569 communicative gaze information (Caruana et al., 2018). The current paradigm would enable this
570 to be directly interrogated to better understand the unique challenges that some autistic
571 individuals face in processing social cues.

572

573 ## Conclusions

574 This study provides evidence that people adaptively *use* and *dismiss* non-communicative gaze
575 information depending on whether it informs the locus of an upcoming joint attention bid. This
576 adaptive ability is not observed to the same extent when people process dynamic arrow stimuli in
577 an analogous task which carefully controls for attention and perceptual task demands. Our
578 findings have important implications for informing theoretical models of gaze processing in
579 dynamic interactive contexts; highlighting the relative advantage people have for extracting
580 others' relevant and predictive non-communicative gaze information. Our experimental approach
581 using carefully controlled yet interactive paradigms also offers a useful tool for advancing the
582 study of social information processing across neurodiverse people.

583

584

585 ## Acknowledgements

586 We would like to take this opportunity to thank Dr. Peter Humburg for his statistical support. We
587 also thank the Technical Research and Innovation Support team, especially Marcus Ockenden
588 and Craig Richardson, for their continued technical support.

589
590
591
592
593
594
595
596

References

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.

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.

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

Burra, N., Mares, I., & Senju, A. (2019). The influence of top-down modulation on the processing of direct gaze. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(5), e1500.

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. M., Woolgar, A., & McArthur, G. (2020). The effect of non-communicative eye movements on joint attention. *Quarterly journal of experimental psychology*, *73*(12), 2389-2402.

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., Inkley, C., Nalepka, P., Kaplan, D. M., & Richardson, M. J. (2021). Gaze facilitates responsivity during hand coordinated joint attention. *Scientific Reports*, *11*(1), 21037.

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.

641 Caruana, N., Stieglitz Ham, H., Brock, J., Woolgar, A., Kloth, N., Palermo, R., & McArthur, G.
642   (2018). Joint attention difficulties in autistic adults: an interactive eye-tracking study.
643   *Autism*, *22*(4), 502-512.
644 Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions*
645   *of the Royal Society of London. Series B: Biological Sciences*, *358*(1430), 315-324.
646 Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & language*, *25*(2),
647   141-168.
648 Conty, L., N'Diaye, K., Tijus, C., & George, N. (2007). When eye creates the contact! ERP
649   evidence for early dissociation between direct and averted gaze motion processing.
650   *Neuropsychologia*, *45*(13), 3024-3037.
651 Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early
652   Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention
653   to Distress. *Developmental Psychology*, *40*(2), 271-283.
654 Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze Cueing of Attention: Visual Attention,
655   Social Cognition, and Individual Differences. *Psychological bulletin*, *133*(4), 694-724.
656 Gale, A., Spratt, G., Chapman, A. J., & Smallbone, A. (1975). EEG correlates of eye contact and
657   interpersonal distance. *Biological psychology*, *3*(4), 237-245.
658 Heft, H. (2003). Affordances, Dynamic Experience, and the Challenge of Reification. *Ecological*
659   *Psychology*, *15*, 149-180.
660 Helminen, T. M., Kaasinen, S. M., & Hietanen, J. K. (2011). Eye contact and arousal: The
661   effects of stimulus duration. *Biological psychology*, *88*(1), 124-130.
662 Hietanen, J. O., Peltola, M. J., & Hietanen, J. K. (2020). Psychophysiological responses to eye
663   contact in a live interaction and in video call. *Psychophysiology*, *57*(6), e13587.
664 Hobson, J. A., & Hobson, R. P. (2007). Identification: The missing link between joint attention
665   and imitation? *Development and psychopathology*, *19*(2), 411-431.
666 Hobson, R. P. (2005). What Puts the Jointness into Joint Attention? *Joint attention:*
667   *Communication and other minds: Issues in philosophy and psychology*, 185.
668 Johnston, J. E., Berry, K. J., & Mielke Jr, P. W. (2006). Measures of effect size for chi-squared
669   and likelihood-ratio goodness-of-fit tests. *Perceptual and motor skills*, *103*(2), 412-414.
670 Kampe, K. K., Frith, C. D., & Frith, U. (2003). "Hey John": signals conveying communicative
671   intention toward the self activate brain regions associated with "mentalizing," regardless
672   of modality. *Journal of Neuroscience*, *23*(12), 5258-5263.
673 Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous child*, *2*(3), 217-250.
674 Kleinke, C. L., & Pohlen, P. D. (1971). Affective and emotional responses as a function of other
675   person's gaze and cooperativeness in a two-person game. *Journal of Personality and*
676   *Social Psychology*, *17*(3), 308.
677 Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, G., & Wykowska, A. (2018). On the role of
678   eye contact in gaze cueing. *Scientific reports*, *8*(1), 17842.
679 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest'. *R package*
680   *version*, *2*(0).
681 Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package "emmeans":
682   Estimated Marginal Means, aka Least-Squares Means. *Compr. R Arch. Netw*, 1-67.
683 Mares, I., Smith, M. L., Johnson, M. H., & Senju, A. (2016). Direct gaze facilitates rapid
684   orienting to faces: Evidence from express saccades and saccadic potentials. *Biological*
685   *psychology*, *121*, 84-90.

686 Mirenda, P. L., Donnellan, A. M., & Yoder, D. E. (1983). Gaze behavior: A new look at an old
687   problem. *Journal of Autism and Developmental Disorders*, *13*(4), 397-409.
688 Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: the role of the
689   dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology*
690   *and Psychiatry*, *44*(6), 793-809.
691 Mundy, P. (2016). *Autism and joint attention: Development, neuroscience, and clinical*
692   *fundamentals*. Guilford Publications.
693 Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical
694   development and autism spectrum disorder. *European Journal of Neuroscience*, *47*(6),
695   497-514.
696 Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition.
697   *Neural Networks*, *23*(8-9), 985-997.
698 Mundy, P., & Neal, A. R. (2000). Neural plasticity, joint attention, and a transactional social-
699   orienting model of autism. In *International review of research in mental retardation*
700   (Vol. 23, pp. 139-168). Elsevier.
701 Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current*
702   *directions in psychological science*, *16*(5), 269-274.
703 Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom
704   presentation in autism. *Development and psychopathology*, *6*(3), 389-401.
705 Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: is it normal? Is it automatic? Is
706   it social? *Development and psychopathology*, *20*(1), 79-97.
707 Nichols, K., & Champness, B. (1971). Eye gaze and the GSR. *Journal of Experimental Social*
708   *Psychology*, *7*(6), 623-626.
709 Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B.
710   (2013). Package 'mass'. *Cran R*.
711 Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., &
712   Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction.
713   *Neuropsychologia*, *44*(5), 718-730.
714 Senju, A., & Johnson, M. H. (2009). The eye contact effect: mechanisms and development.
715   *Trends in cognitive sciences*, *13*(3), 127-134.
716 Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Harvard
717   University Press Cambridge, MA.
718 Tomasello, M. (1995). Joint attention as social cognition. *Joint attention: Its origins and role in*
719   *development*, *103130*.
720 Trevisan, D. A., Hoskyn, M., & Birmingham, E. (2018). Facial expression production in autism:
721   A meta-analysis. *Autism Research*, *11*(12), 1586-1601.
722 Wilson, D., & Sperber, D. (2002). Relevance Theory. In L. R. Horn & G. Ward (Eds.),
723   *Handbook of Pragmatics*. Blackwell.
724 Xu, S., Zhang, S., & Geng, H. (2018). The effect of eye contact is contingent on visual
725   awareness. *Frontiers in Psychology*, *9*, 93.
726 Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005).
727   Behavioral manifestations of autism in the first year of life. *International journal of*
728   *developmental neuroscience*, *23*(2-3), 143-152.

729

Figure 1

Stimuli used in the interactive joint attention task.

Figure shows 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.

Figure 2

Schematic representation of trial sequence and central stimulus used by condition.

(A) Trial sequence for social Eyes stimuli for each of the three context conditions (NoSearch, Random Search, Predictive Search). The eye icon in the figure represents the fixation location made by the participants at each point in the trial and was not visible to them. Saccadic reaction time (SRT) was measured as the latency between cue presentation and the onset of responsive saccades. (B) An example of the central stimulus used in each condition, representing the analogous non-social stimulus used in the arrow condition for eyes closed, eye contact, and eyes averted. The background stimulus and the timing of stimulus presentation were identical across stimulus conditions (Eyes, Arrows).

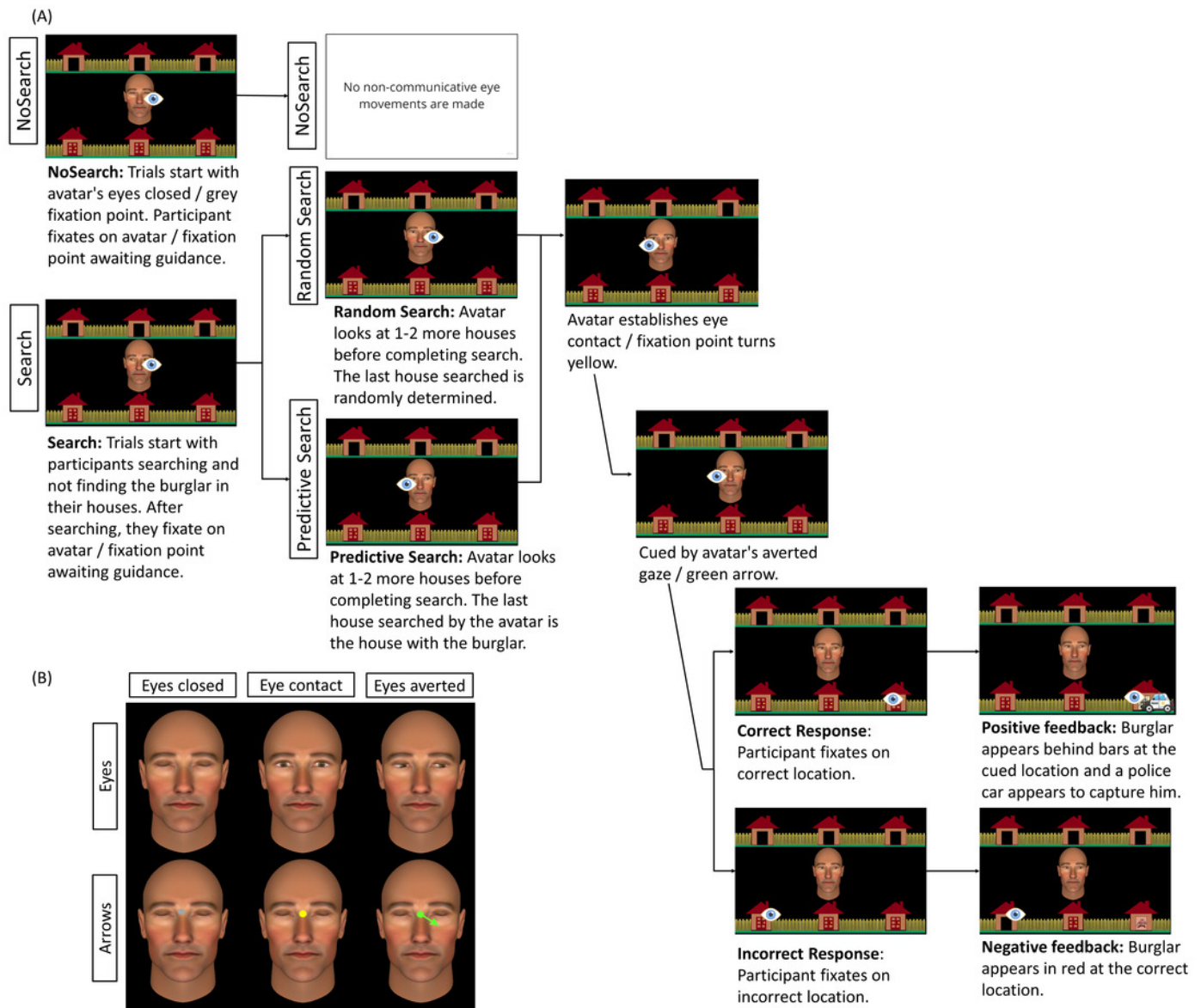



Figure 3

Raincloud plot with individual data points illustrating the proportion of correct trials by context (Random Search, No Search, Predictive 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 (*** $p < .001$).

Accuracy

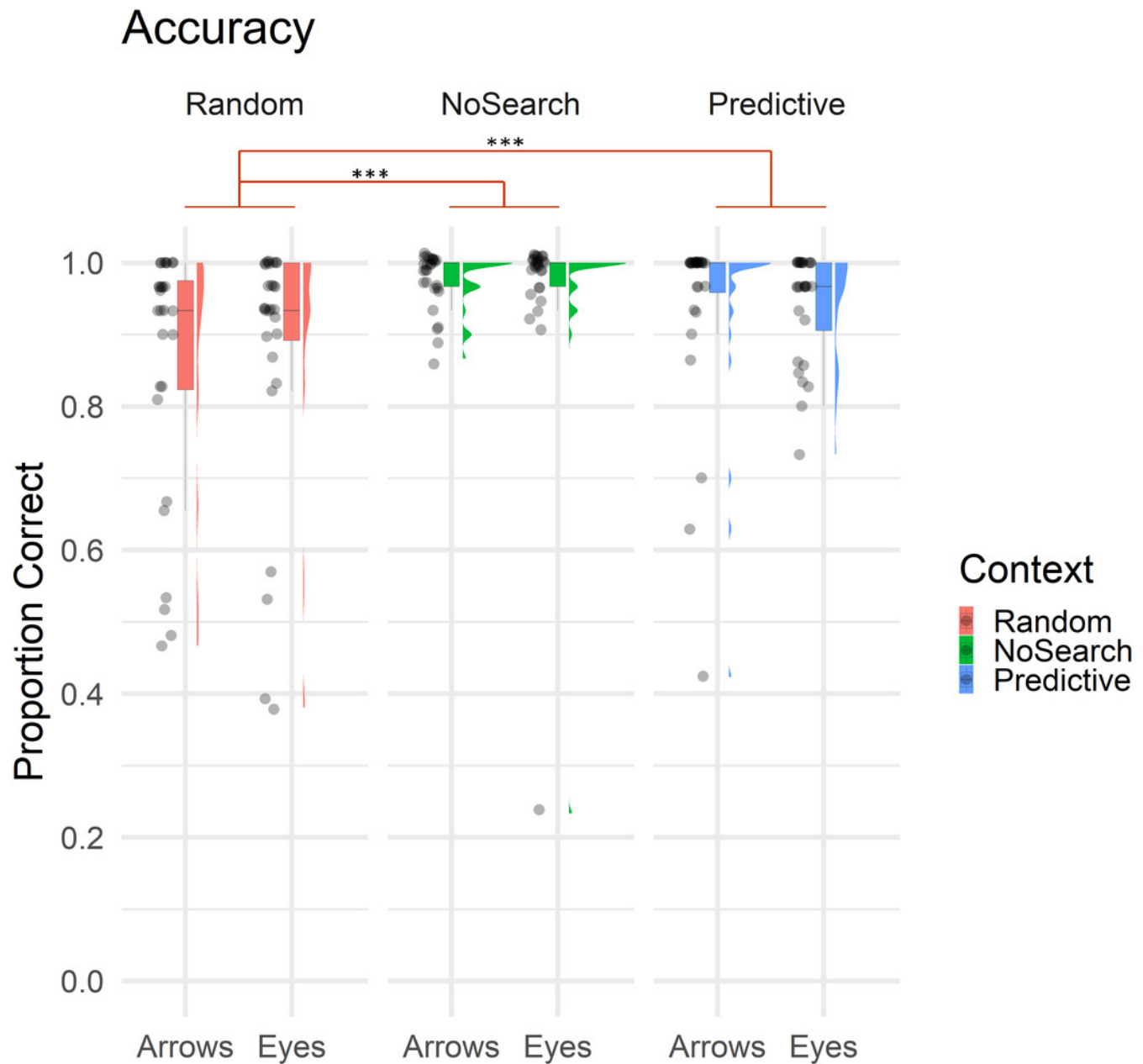
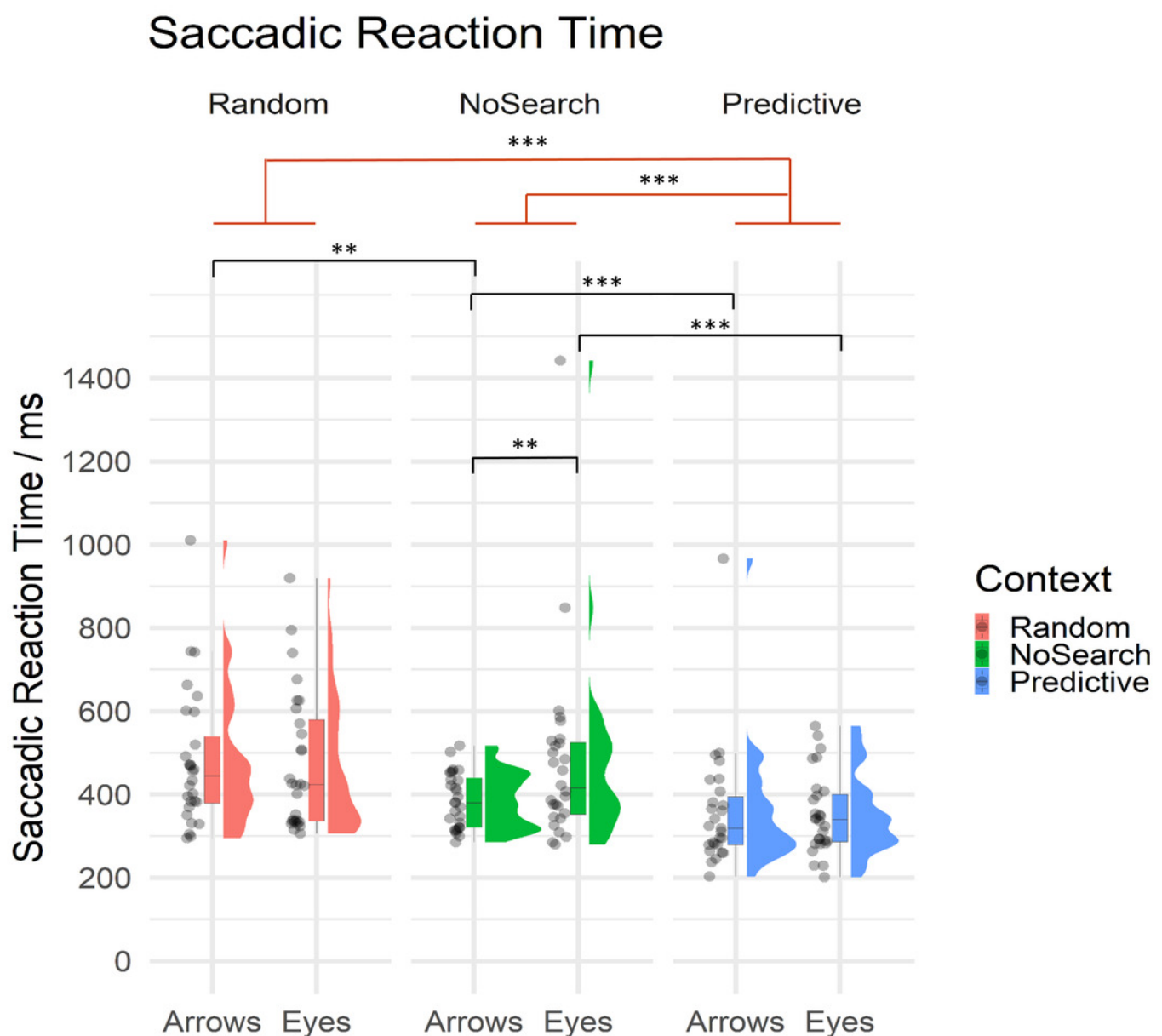



Figure 4

Raincloud plot 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).

Estimated fixed effect parameters for Accuracy.

1 **Table 1:**

2 **Estimated fixed effect parameters for Accuracy.**

Fixed effect	β-coefficient	Standard Error (SE)	z-value	p-value
Context				
NoSearch-Random	1.857	0.348	5.335	< .000***
Predictive-NoSearch	-0.657	0.460	-1.428	0.153
Predictive-Random	-2.400	0.642	-3.738	< .000***
Stimulus				
Arrows-Eyes	-0.057	0.302	-0.190	0.849
Stimulus*Context				
NoSearch-Random	-0.828	0.453	-1.828	0.068
Predictive-NoSearch	-0.066	0.499	-0.133	0.894
Random-Predictive	-0.894	0.383	-2.338	0.019*
Follow-up comparisons				
Random-Predictive (Arrows)	-1.647	0.382	-4.311	<.000***[a]
Random-Predictive (Eyes)	-0.753	0.365	-2.062	0.039[a]

3 *Note. *p < .05, ** p < .01, *** p < .001, [a] Corrected p-values using a FDR correction for*

4 *multiple comparisons.*

Means (M) and standard deviations (SD) of SRT by Condition.

1 **Table 2:**

2 Means and standard deviations (SD) of SRT by Condition

Condition	Random Arrows	NoSearch Arrows	Predictive Arrows	Random Eyes	NoSearch Eyes	Predictive Eyes
M (SD)	470.90 ms	382.29 ms	338.65 ms	474.75 ms	445.04 ms	339.44 ms
	(351.27)	(227.21)	(253.67)	(344.56)	(281.94)	(196.38)

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

Estimated fixed effect parameters for SRT.

1 **Table 3:**

2 **Estimated fixed effect parameters for SRT.**

Effect	β-coefficient	Standard Error (SE)	t-ratio	p-value
Context				
NoSearch-Random	0.149	0.076	1.978	0.058
Predictive-NoSearch	0.635	0.108	5.871	< .000***
Predictive-Random	-1.568	0.179	-8.783	< .000***
Stimulus				
Arrows-Eyes	-0.156	0.077	-2.016	0.050*
Stimulus*Context				
NoSearch-Random	-0.273	0.067	-4.075	< .000***
Predictive-NoSearch	0.201	0.069	2.910	0.004**
Random-Predictive	-0.072	0.072	-1.002	0.316
Follow-up comparisons				
NoSearch-Random (Arrows)	0.286	0.083	3.446	0.002**[a]
NoSearch-Random (Eyes)	0.013	0.083	0.158	0.875[a]
Predictive-NoSearch (Arrows)	0.534	0.114	4.702	< .000***[a]
Predictive-NoSearch (Eyes)	0.735	0.113	6.477	< .000***[a]
Arrows-Eyes (NoSearch)	0.314	0.085	3.706	.001**[a]
Arrows-Eyes (Random)	0.041	0.088	0.464	0.752[a]
Arrows-Eyes (Predictive)	0.113	0.089	1.265	0.294[a]

3 *Note. *p < .05, ** p < .01, *** p < .001, [a] Corrected p-values using a FDR correction for*

4 *multiple comparisons.*