1

2

3

4   **Awareness of gaze behaviour is limited: Insights from a novel tracking paradigm**

5

6   Avery Hannah Chua[1] and Anna Kosovicheva[1]

7   [1] Department of Psychology, University of Toronto Mississauga

8

9   **Author Note**

10   Avery Hannah Chua https://orcid.org/0009-0000-6817-6721

11   Anna Kosovicheva https://orcid.org/0000-0002-5219-3006

12   We have no conflict of interest to disclose. Correspondence concerning this article should

13   be addressed to Avery Hannah Chua, chuaa1216@uerm.edu.ph)

Abstract

14

15   When asked where they have previously looked, people rarely report their visual

16   behaviour correctly. However, previous tasks have probed participants' awareness of their gaze

17   behaviour through explicit measures, like recognition, and it remains unclear whether poor

18   awareness is also seen in implicit measures. We investigated this with a novel tracking paradigm

19   in which participants first completed a visual search task while their eye movements were

20   recorded. Next, participants completed a tracking task where they followed a moving red dot on

21   the screen with their eyes. In Experiment 1, the dot replayed either their own previously recorded

22   gaze position or that of another participant. We measured tracking performance by cross-

23   correlating the previously recorded gaze position with the tracked positions. Participants were

24   not significantly faster or more accurate in tracking their own eye movements compared to

25   another participant's. Experiment 2 tested this with a more extreme manipulation, in which

26   participants tracked either unaltered or temporally reversed sequences of their own eye

27   movements, which resulted in higher accuracy in the forward compared to the reversed

28   condition. Finally, Experiment 3 examined whether performance would decrease when

29   participants tracked a participant with very dissimilar gaze behaviour from their own. Tracking

30   performance was similar when tracking their own scanpath compared to a dissimilar participant,

31   but participants' tracking latencies were shorter when tracking a participant with consistent

32   scanning behaviour. Overall, our results suggest that though awareness of one's eye movements

33   is generally poor, extreme manipulations, like temporally reversing one's eye movements, can

34   influence tracking performance.

35   *Keywords:* awareness, eye movements, gaze behaviour, individual differences

36   Eye movements are an essential part of how we gather information from the world. Every

37   day, when we look for our house keys or try to spot a friend in the crowd, we make directed eye

38   movements towards objects of interest across the visual field. Previous research has shown that

39   the planning and programming of eye movements is reliant on a combination of top-down

40   influences like strategy (Boot et al., 2009; Clarke et al., n.d.) and bottom-up factors like saliency (Itti

41   & Koch, 2000; Theeuwes et al., 1998, 1999). Moreover, these factors are often highly idiosyncratic

42   and are a product of the limitations of the visual system as well as cognitive variables, which

43   translates into a vast array of individual differences in gaze behaviour across different tasks

44   (Andrews & Coppola, 1999; Clarke et al., n.d.; Henderson & Luke, 2014; Rayner et al., 2007; Veríssimo et

45   al., 2021).

46   In most everyday tasks, it may not be necessary for us to know where we last looked, or

47   to maintain a robust and lasting representation of this information. This makes intuitive sense,

48   given that we make two to four saccades per second, each separated by about 200-300 ms

49   (Carpenter, 1988). In other words, storing detailed information of our own gaze behaviour is

50   computationally costly for the brain and may be unnecessary. However, there are cases where it

51   may be beneficial to have some knowledge of our own gaze behaviour. For example, when we

52   search for items, knowledge of where we last looked may improve efficiency and prevent

53   revisitations to previously examined locations (Klein, 1988; Posner & Cohen, 1984). This may

54   also be crucial to accurately gauge when to terminate search, especially when the target is absent

55   (Kok et al., 2017; Võ et al., 2016). Put more simply, how do we know that we have "looked

56   everywhere"? Lastly, with increased interest in eye-tracking as a way to train gaze behaviour in

57   real-world tasks, like radiology or airport baggage screening (e.g., Ashraf et al., 2018; Sadasivan et

58   al., 2005; van Gog & Jarodzka, 2013), it would be useful to know if we are able to monitor our gaze

59   behaviour for learning and self-regulation, since monitoring one's own behaviour has been

60   shown to be important for learning in non-visual domains (Greene & Azevedo, 2010; Nelson, 1990).

61   Several studies have examined the extent of our awareness of our gaze behaviour, finding

62   that we generally have somewhat limited knowledge of where we have previously looked.

63   Foulsham and Kingstone (2013) were the first to explicitly ask if people could identify their own

64   fixation pattern compared to one of the following foils: random fixations, fixations from another

65   image, and fixations from another person. Performance was the lowest when participants were

66   tasked with distinguishing their own fixations from another person's, but remained above chance

67   levels in all conditions. In a separate study, when asked to replicate their previous fixations using

68   mouse clicks, participants missed fixations and made false reports (Marti et al., 2015), though for

69   the most part, they were able to report their gaze behaviour correctly.

70   However, there are a number of challenges in probing participants' knowledge of their

71   own gaze behaviour. For example, participants may develop particular gaze strategies if they

72   know that they will be asked about where they've looked, or rely on other sources of

73   information. For example, participants may infer that they made more fixations when the trial

74   was longer (Marti et al., 2015). Participants might also remember the objects they fixated on

75   (Foulsham & Kingstone, 2013) or the target object being searched for (Clarke et al., 2017), but not

76   necessarily where they've looked. Previous studies controlling for these variables show reduced

77   ability to report on one's own gaze behaviour. For example, Võ et al. (2016) also used mouse

78   clicks, but only probed one-fourth of the total trials, to prevent people from adopting a strategy to

79   boost their performance. They found that people were no better at distinguishing between their

80   own fixations from another person's. In another study, Clarke et al. (2017), separately analyzed

81   participants' memory for objects within a scene compared to memory for objects that had been

82   fixated. Participants had difficulty distinguishing objects they had fixated from objects they had

83   not. To add to this, other research has shown that even when using a gaze-contingent display,

84   which showed participants their own eye movements in real-time, participants' ability to identify

85   their own eye movements remained incredibly poor (Kok et al., 2017). Together, these studies

86   report that our awareness of our own gaze behaviour is very limited.

87   These studies also highlight the importance of controlling for extraneous variables,

88   including strategy or other sources of information. Explicit probes, like forced choice or free

89   report tasks, may not fully capture participants' knowledge of their own gaze behaviour.

90   Moreover, since the majority of these are more likely to test memory, rather than awareness, they

91   may be prone to serial position effects, limited working memory capacity, and poor

92   metacognition.

93   Another possibility is that we retain some knowledge of our typical gaze behaviour, but it

94   is not accessible through explicit recall or recognition. To investigate whether this is the case, we

95   designed a novel tracking paradigm as an indirect probe of whether participants have some

96   information about their typical gaze behaviour. In this paradigm, we first recorded participants'

97   eye movements while they completed a visual search task. Next, participants reproduced their

98   previous eye movements that were played back to them by following a dot that moved around

99   the screen with their eyes. The underlying logic is that, if participants have some knowledge of

100 where they tend to look, whether it is accessible to them or not, they should be faster and more

101 accurate in reproducing that same behaviour compared to another individual's sequence or an

102 altered scanpath.

103   This task design was partly based on previous work showing that dynamic displays boost

104 recognition performance (Van Wermeskerken et al., 2018). In this previous study, participants were

105   asked to discriminate their own eye movements from someone else's in a forced choice task.

106   Recognition performance was better in this task when viewing dynamic replays compared to

107   static images, likely due to the presence of both spatial and temporal information. Similarly, our

108   dynamic displays preserved the spatial and temporal information in the original gaze recording.

109   In addition to this, by using an implicit measure, we did not probe participants' memory for

110   where they looked on a particular trial. Instead, the measure is intended to capture whether

111   participants have some information about their typical gaze behaviour within a search array. In

112   addition, we used a simple visual search display (a search for T shapes among distractors) to

113   minimize the impact of scene content on performance.

114   **Experiment 1**

115   In Experiment 1, we tested our novel tracking task, which involved dynamic playback of

116   previously recorded eye movements. As shown in Figure 1A, participants completed two tasks: a

117   search task and a tracking task. In the search task, on each trial, participants searched for a

118   perfect "T" shape among near-"T" shapes and "L"-shaped distractors (Figure 1B). In the tracking

119   task, accuracy and latency in tracking performance were measured by cross-correlating the

120   previously recorded gaze position with the tracked gaze position (Figure 2). To identify whether

121   people are able to track their own eye movements better than those of another person, we

122   compared tracking accuracy and latency for previously recorded eye movements from the same

123   participant versus the previous participant. In addition, we played the gaze positions back either

124   on the same background that they were recorded from, or on a blank background, to test the

125   prediction that tracking performance would improve when the set of possible gaze locations is

126   constrained by the visual information in the search stimulus array (i.e., display). We compared

127   tracking accuracy in a 2×2 design (Figure 1C): replaying eye movements that were the

128   participant's own versus another observer's (Self versus Other), and superimposing the replayed

129   position on the original stimulus array or not (Background versus No Background).

<div align="center">

Methods

</div>

131   **Participants**

132   Thirty participants were recruited from the paid participant pool at the University of

133   Toronto Mississauga. Following exclusions (see *Exclusion Criteria*), the final sample consisted

134   of twenty-four participants (age: $M = 20$; 6 males, 16 females, and 2 unreported). Two

135   participants were excluded due to a high proportion of lost eye-tracking samples (greater than

136   20%), while another was excluded due to low accuracy in the search task (not significantly better

137   than 50%, based on a binomial test), and three did not finish the study and their data were

138   discarded. All participants had normal or corrected-to-normal vision, defined as 20/25 or better

139   in each eye using an ETDRS Near Vision chart. Participants provided informed consent prior to

140   participating in the study, and all study procedures were approved by the Research Ethics Board

141   at the University of Toronto. Two participants were compensated with course credit in

142   accordance with their request, while the rest were compensated at $15 CAD/hour. On average,

143   the task took between 75 to 90 minutes.

144   **Material and Stimuli**

145   *Apparatus.* Participants maintained a stable viewing distance of 57 cm from a gamma-

146   corrected LG UltraGear LCD screen (2560 × 1440 pixels; 60 Hz refresh rate), such that the full

147   display spanned approximately 55.3º × 32.8º. All visual stimuli were programmed using

148   MATLAB and the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

149   The experiment was run on an Apple Mac Pro running MacOS 10.14.6 and a NVidia GTX 680

150   4GB graphics card. Throughout the experiment, eye movements were recorded using an Eyelink

151   1000 Tower Mount tracker, which sampled at 1000 Hz. All eye movements were recorded from

152   the right eye. A chin and head rest were used to minimize movements during the experiment.

153   Calibration was performed prior to the experiment and at the half-way point. Drift correction was

154   performed every 40 trials, in which participants were required to fixate on a central dot prior to

155   resuming the task and the eye-tracking recording. Additional calibration was performed when

156   needed.

157   *Search stimulus array.* Each stimulus array (Figure 1B) consisted of 25 varying shapes

158   assigned to randomly selected positions in an invisible 8-row × 11-column screen-centred grid

159   that spanned 15º × 20.6º on a gray background (96.1 cd/m^2). On each trial, random horizontal

160   and vertical spatial jitters were added to the center positions of each of the 8 × 11 grid positions,

161   ranging from -0.59º to +0.59º to prevent overlap in the stimuli. The shapes all consisted of one

162   horizontal and one vertical line segment, joined at right angles to form a continuum of shapes

163   ranging from a perfect T to an L. The line segments were 0.23º thick with each bar of the 'T' or

164   'L'-variant shapes set to 1.29º in length. Search targets were perfect "T" shapes, such that the

165   vertical bar perfectly bisected the horizontal bar. Distractors were non-T stimuli that randomly

166   varied in the offset of their horizontal and vertical bars. To clearly differentiate the non-T

167   distractors from the T-shapes, the vertical segment was offset from the midline of the horizontal

168   segment by at least 0.12º, and up to a maximum of 0.52º (making a perfect "L" shape). Likewise,

169   the stimuli, including the target T, randomly varied in their orientation (selected from 0°, 90°,

170   180°, or 270°) and greyscale intensity value, which ranged from 38.4 to 57.6 cd/m^2 (different

171   shades of dark grey). Each stimulus array was randomly generated, such that a new array was

172   presented on each trial and for each participant.

173   *Tracking conditions.* In the second half of the experiment, participants completed a

174   tracking task (Figure 1C) in which they followed a moving red dot on the screen (0.9º diameter)

175   with their eyes. On each trial, the dot's position matched either their own previously recorded

176   gaze position or that of the previous participant (Self versus Other conditions, respectively).

177   Gaze positions were replayed from a random subset of 50% of the trials corresponding to these

178   conditions, and in a random order, such that no sequence was replayed more than once. Positions

179   were downsampled from the original 1000 Hz recording to the display frame rate (60 Hz) to

180   match the original speed of the recorded eye movements. Furthermore, replayed eye movements

181   were either superimposed on the same stimulus array displayed during the search task that they

182   were originally recorded from, or on a plain grey background (Background versus No
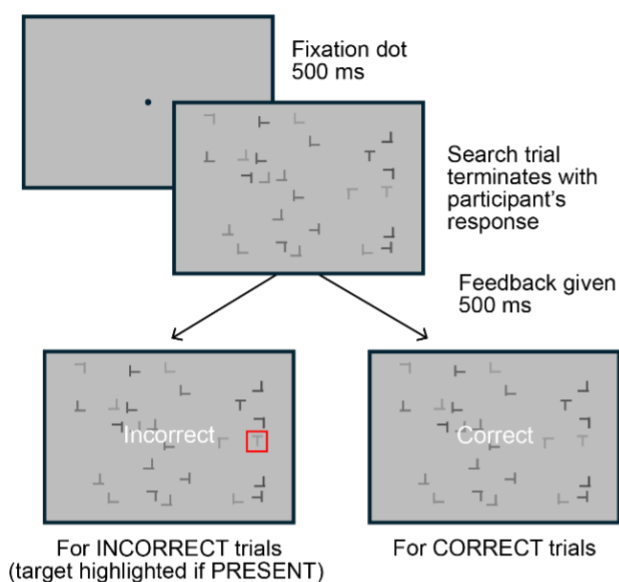
183   Background).

184   In both conditions, the dot was visible for only 50% of the time and randomly

185   disappeared for brief segments of time. Pilot experiments showed near-ceiling tracking

186   performance across all conditions when the dot was visible for the full duration, and this was

187   mitigated by the visibility manipulation. As shown in Figure 1D, the visibility manipulation was

188   introduced by splitting the gaze position time series into segments that were 20 frames (333 ms)

189   long, except for the last segment, which spanned any remaining frames. A random 50% of the

190   segments were selected to have the dot visible, while the other 50% were selected to have the dot

191   invisible. The result is that participants saw a red dot moving overlaid on the display elements

192   (stimulus array present) or on a grey background (stimulus array absent) and would

193   unpredictably disappear and reappear for brief periods of time. This duration of each segment

194   was chosen to capture the length of approximately one fixation. Participants completed 20

195   practice trials before another 160 experimental trials (Figure 1A). In total, the 2×2 experimental

196   design consisted of 40 trials per unique condition of the tracking task (e.g., 40 Self-Background

197   Present trials; 40 Other-Background Present trials, etc.), with conditions randomly interleaved
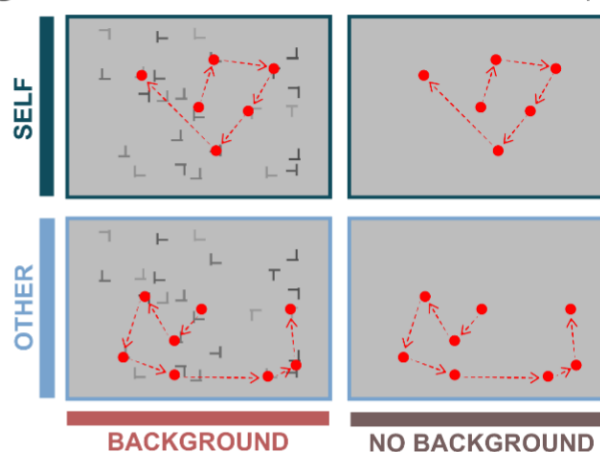
198   across trials.

199

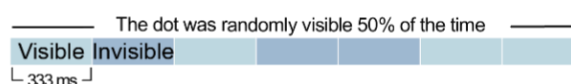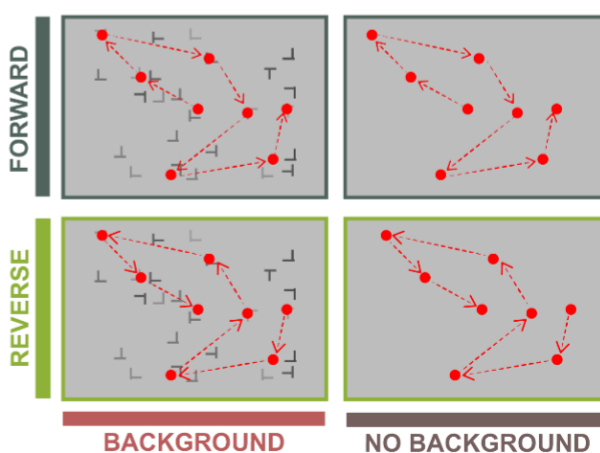**Figure 1. Design and procedure for Experiments 1 and 2. (A)** *Schematic of trial order.* Participants completed 20 practice trials followed by 160 trials of the experiment proper for both the search and tracking tasks, with an option for breaks after every 40 trials. In between the two tasks, a longer break occurred, in which we processed the eye-tracking data and prepared it for the tracking task. **(B)** *Trial sequence for search task.* Participants were presented with a fixation dot for 500 ms prior to the search stimulus array. Search trials terminated upon key response (left arrow for "absent" and right arrow for "present"), followed by a feedback screen for 500 ms. **(C)** *Conditions for the tracking task, Experiment 1.* Participants were either presented with replayed eye movements from themselves (Self) or from the previous participant (Other), superimposed on top of the original stimulus array (Background) or on a grey background (No Background). Conditions were randomly interleaved. **(D)** *Visibility manipulation.* To reduce ceiling effects on tracking performance, the dot was visible for 50% of the time, which was done by dividing the trial length into segments (333 ms each), and randomly selecting 50% of these segments to have the dot visible. **(E)** *Spatial arrangements for Experiments 2 and 3.* Item locations in these experiments were selected from a set of ten possible spatial arrangement (layout). Each iteration of the same spatial arrangement was different, with a different target location (if present), orientation, and color intensity. In short, only the spatial arrangement remained the same across all iterations. The red circles in the third panel emphasize that across all iterations, only the locations of these items remained constant. **(F)** *Conditions for tracking task, Experiment 2.* Participants were either presented with their own, unaltered eye movements (Forward) or a temporally reversed sequence (Reverse), superimposed on top of the original stimulus array (Background) or on a grey background (No Background).

Procedure

Search task. As shown in Figure 1B, in the first half of the experiment, participants

completed a visual search task consisting of L-shapes and T-shapes, where they were instructed

to identify the presence of a perfect T among 25 other distractors. Trials terminated with the

participant's response via a key press (left arrow key for 'absent' and right arrow key for

'present'). In total, participants completed 20 practice and 160 experimental search trials, with

half of the trials containing a target (50% target present). Participants were given the opportunity

to take a break every 40 trials. Each trial sequence started with a fixation dot (a small black dot

in the center of the screen, 0.46° diameter) presented for 500 ms, followed by the visual search

display that terminated upon participant response. Feedback regarding accuracy was then shown

for 500 ms (the text "Correct" vs "Incorrect" at the center of the display, see Figure 1B), with a

red square outlining the perfect T when participants responded 'absent' on target-present trials.

234   Lastly, participants were not informed of the second task to prevent deliberate gaze strategies

235   that would allow them to recall their previous eye movements. Once the search task was

236   completed, eye-tracking data were processed and cleaned in between the two experimental

237   blocks. Eye blinks and other missing samples were replaced by interpolated data (using the

238   nearest-neighbour method), for the duration of the eye blink as well as the 70 ms before and after

239   the blink. This was done to maintain experimental control over the visibility of the dot during the

240   tracking task (i.e., such that it would be visible for 50% of the time and not more).

241   *Tracking task.* On continuing to the second experimental block, participants were

242   presented with previously recorded gaze positions represented as a red dot (Figure 1C). On each

243   trial, participants first saw a fixation dot (identical to the one present during the *search task*)

244   presented for 500 ms, followed by the red dot, which represented either their own previously

245   recorded gaze positions or those of the previous participant (N-1, see *Exclusion Criteria* for how

246   we replaced participants and ensured continual staggering of the data presented). Replayed

247   recordings were matched onto the original stimulus array that the gaze positions were recorded

248   from during Background-Present trials, while in Background-Absent trials, participants only saw

249   the red dot on a plain grey background. This was to determine whether having a stimulus array

250   present would facilitate tracking performance. On each trial in the tracking task, participants

251   were then instructed to follow the red dot as accurately as possible. If invisible, they were told to

252   move their eyes to where they believed the dot would appear next, until the red dot appeared

253   again. At the end of each trial, participants manually initiated the next trial by pressing the

254   spacebar. Importantly, participants were not instructed to search for any target, but were simply

255   told to follow the dot with their eyes. Participants were not informed of the condition before each

256   trial, nor were they informed of the different unique conditions in the *tracking task*. As before,

257   participants had the option for breaks after every 40 trials, with a drift correction prior to

258   resuming the task.

259   **Analysis**

260   *Exclusion Criteria.* Data were assessed for overall quality based on the following

261   preregistered criteria: (1) *Search performance*. Participant data were excluded and replaced if

262   their overall search accuracy was not significantly better than chance (50% accuracy) based on a

263   binomial test, or if their median reaction times were excessively fast (< 3.5 s, based on pilot data;

264   see *Participants*). As these measurements only apply to the search task, these exclusions were

265   based on the first experimental block only. Given our Self versus Other manipulation, we

266   ensured that replacement participants viewed the gaze paths from the participant prior to the

267   excluded participant in the 'Other' condition. For example, if participant N was following X, but

268   N was excluded, then the replacement Y participant will follow X's eye movements as well. (2)

269   *Eye tracking quality.* Participants with 20% or more missing or unusable eye-tracking samples

270   during the first experimental block were excluded. Lost or unusable eye tracking are gaze

271   positions undetected or reported outside the bounds of the tracking region (a 19.7º × 25.1º area

272   centred on the search display).

273   *Tracking measures.* Prior to analysis, eye-tracking data were processed and cleaned using

274   the same method for processing the gaze data for the tracking task (i.e., removing missing

275   samples and interpolating them using the nearest-neighbour method). As shown in Figure 2,

276   tracking accuracy and latency were calculated by cross-correlating the original gaze position in

277   the search task with the corresponding gaze positions in the tracking task. For the search

278   scanpath, we used all samples in the original recording, without any breaks or gaps. In other

279   words, these included gaze samples from the blank segments, not shown to the participant in the

280   tracking task. To calculate the cross-correlograms (CCGs), we removed the first 500 ms of the

281   trial recording to account for participants orienting to the dot's position from the fixation dot.

282   Moreover, we only analyzed up to 4000 ms of each trial to ensure the same length of data across

283   our analysis. As participants varied in their reaction times, and therefore the trial duration, this

284   was done to match the number of samples used for calculating the CCG between the Self and

285   Other conditions (though we note that we found similar results when using all samples in each

286   trial; see Supplemental Figure 1). Trials that were shorter than 4000 ms were excluded from the

287   analysis. From here, separate cross-correlations were calculated from the replayed x- and y- gaze

288   positions (horizontal and vertical positions, respectively) collected during the first experimental

289   block (visual search task) and the recorded x- and y- gaze positions from the second

290   experimental block (tracking task).

291   Tracking accuracy was calculated from the resulting x- and y-peak correlation values

292   from each trial. These values were Fisher Z-transformed prior to averaging across trials, and then

293   averaged across x- and y-values for an overall tracking accuracy measure. This provided a

294   measure of tracking accuracy, accounting for lags. Similarly, we extracted the delay

295   corresponding to the peak in the cross-correlation as the tracking latency measure. This was first

296   calculated from each trial, and then averaged across all trials within a condition, and then

297   averaged across the x- and y- coordinates. Prior to averaging either the peak or latency values,

298   we removed outliers using the interquartile range method (IQR), such that individual trials that

299   were more than 1.5 interquartile ranges above the upper quartile or below the lower quartile were

300   removed. This resulted in the removal of 4.35% of peak or latency values on average across

301   participants from all three experiments. Both tracking accuracy and latency were then compared

302   across the four conditions (Background-Self, Background-Other, No Background-Self, No

303   Background-Other) using a 2×2 repeated-measures ANOVA.
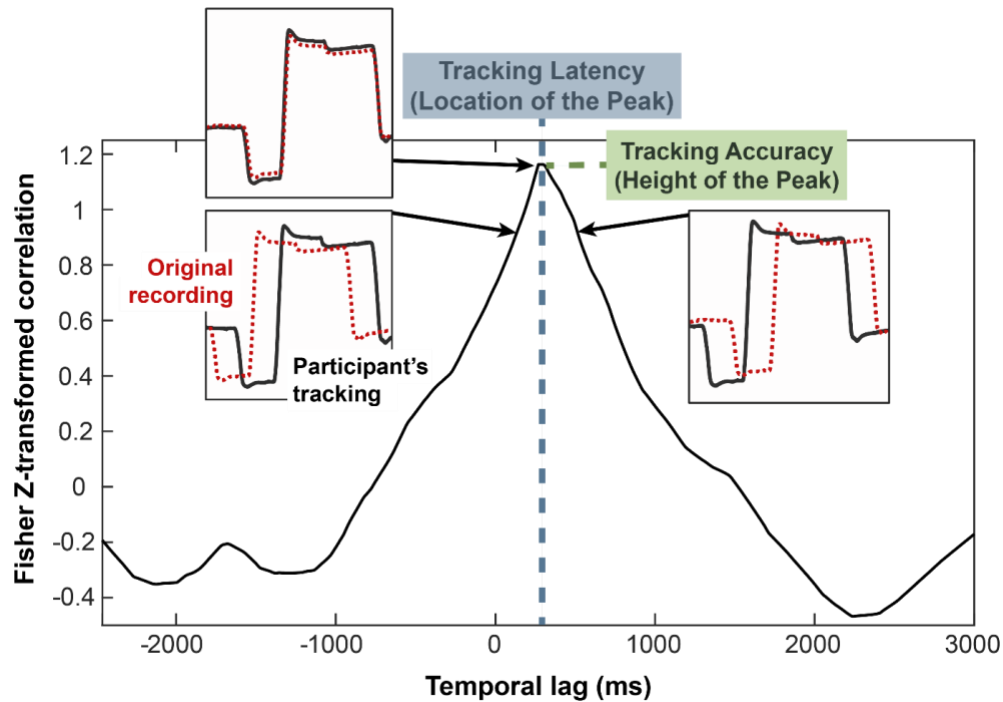

304

305   **Figure 2. Schematic of analysis using cross-correlation.** Gaze positions collected from the first
306   experimental block (red dotted line, showing horizontal position over time, in inset panel, replayed as a red
307   dot) were cross-correlated with the gaze positions recorded (solid black line in inset panel) during the
308   tracking task to calculate our dependent variables. The graph shows a sample cross-correlogram (CCG),
309   which illustrates the correlation between original and tracked gaze positions as a function of time lag.
310   Positive lags on the x-axis correspond to the original recording being shifted later, while negative lags
311   correspond to the original recording being shifted earlier. The height of the peak represents the tracking
312   accuracy, and the location of the peak represents the temporal lag between the two time series. A lag at a
313   positive value indicates that the two time series are most correlated when the original recording is shifted
314   later.

315   **Results**

316   **Tracking accuracy.** Following our analysis described above, we calculated the average

317   tracking accuracy for each condition (Figure 3A). A 2 (source of eye movements: Self versus

318   Other) × 2 (stimulus array: Background versus No Background) repeated-measures ANOVA

319   revealed no significant main effect of eye movement source, $F(1,23) = 0.03$, $p = .875$.

320   Participants' tracking accuracy was no different when the dot replayed eye movements from

321   themselves (Self; $M = 0.94$, $SD = 0.10$) compared to another participant's eye movements (Other;

322    $M = 0.94$, $SD = 0.10$). Likewise, the presence or absence of the stimulus array did not affect

323   tracking accuracy, $F(1,23) = 0.5121$, $p = .481$, ($M = 0.95$, $SD = 0.11$ versus $M = 0.94$, $SD = 0.10$).

324   There was also no significant interaction, $F(1,23) = 0.01$, $p = .916$. On average, participants were

325   not significantly better at tracking their own eye movements compared to another's, nor did the

326   presence of the original stimulus array influence tracking accuracy.

327   **Tracking latency.** A separate 2×2 repeated-measures ANOVA revealed a significant

328   main effect of stimulus array presence ($F(1,23) = 44.25$, $p < .001$, $\eta^2_p = 0.658$) on tracking

329   latency. As shown in Figure 3B, on average, participants had a lower latency (i.e., faster

330   tracking), when the original stimulus array was present ($M = 207.62$ ms, $SD = 43.68$ ms)

331   compared to when it was absent ($M = 241.83$ ms, $SD = 35.19$ ms). However, no main effect of

332   eye movement source was found ($F(1,23) = 0.33$, $p = .573$) meaning that, on average, people did

333   not have shorter latencies when the gaze positions presented came from themselves (Self; $M =$

334   221.95 ms $SD = 41.95$ ms) or from another participant (Other; $M = 227.50$ ms, $SD = 44.37$ ms).

335   There was also no significant interaction, $F(1,23) = 0.40$, $p = .533$. In sum, these results suggest

336   a latency benefit when the original stimulus array was present, regardless of the source of the eye
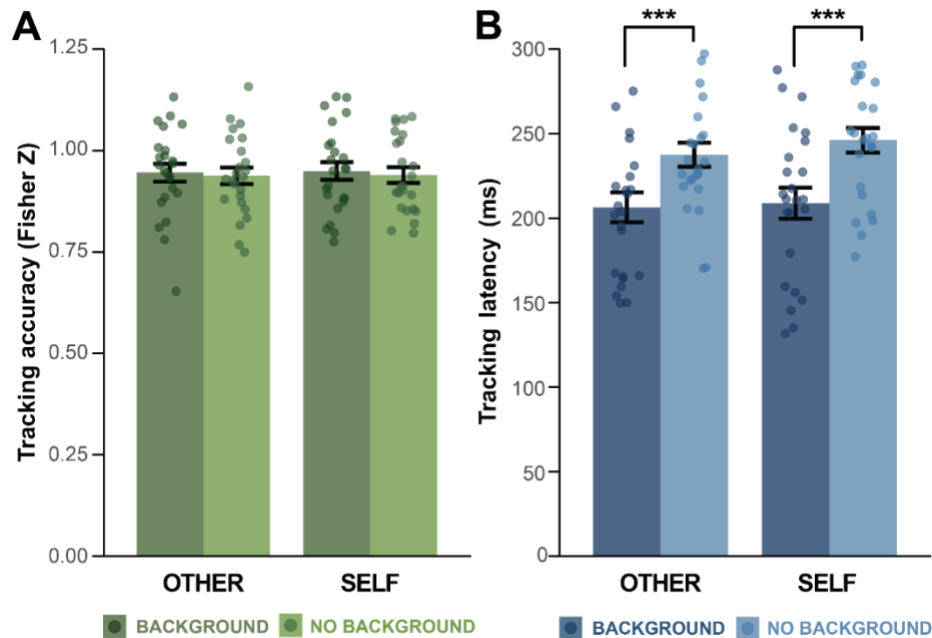
337   movements.

Figure 3. Tracking performance in Experiment 1. (A) Average tracking accuracy in each condition, calculated as the peak of the cross-correlation (Fisher Z-transformed correlation value). (B) Average tracking latency in each condition, calculated as the lag corresponding to the peak accuracy (i.e., the point on the x-axis of the CCG corresponding to the peak). Symbols represent individual participants, and error bars represent ± 1 standard error of the mean (SEM). Asterisks represent results of the 2×2 repeated-measures ANOVA, ***$p < .001$.

Discussion

Experiment 1 showed no difference in tracking accuracy or latency between the Self and Other conditions, which may be consistent with several interpretations. Firstly, the lack of a difference may suggest that the idiosyncrasies in gaze behaviour are too small to detect with this method. For example, this could result from inconsistent gaze patterns. When people employ random eye movement sequences across randomly generated visual search arrays, a lack of a consistent difference between individuals in scanning behaviour would produce comparable tracking performance for their own eye movements versus others'. Similarly, these results may also stem from the staggered nature of the Other condition. To elaborate, the eye movements for the Other condition always came from the person before the current participant. This method

355   may pair people with similar gaze behaviour by random chance, which could make it unlikely or

356   difficult to produce a difference in tracking accuracy between their own gaze patterns and those

357   of another participant. However, given that each trial was a unique search array with a distinct

358   target location and set of distractors, we are limited in our ability to answer this question. Finally,

359   despite the limitations mentioned above, our results may be a true product of poor knowledge of

360   one's own gaze behaviour, which would be consistent with previous literature (Clarke et al.,

361   2017; Kok et al., 2017; Marti et al., 2015; Võ et al., 2016). The results of Experiment 1 alone are

362   not enough to establish whether or not participants retain information about their own typical

363   gaze behaviour.

364   Unsurprisingly, Experiment 1 showed a latency benefit for the Background Present

365   compared to the Background Absent condition. Participants' tracking latency was lower when the

366   original stimulus array was present. Having this stimulus array of 25 items visible during the

367   replay phase of the experiment constrained the set of possible locations where the dot would

368   likely appear next. In contrast, when the stimulus array was absent, the number of possible

369   locations is much higher. By reducing the number of possible saccade locations, we also reduce

370   the computational demand on saccadic planning and programming (Abrams & Jonides, 1988), as

371   evident by the latency difference of about 40 ms. Overall, Experiment 1 served as a proof of

372   concept for our novel tracking task and its potential as an implicit probe for awareness.

Experiment 2

374   Experiment 1 demonstrated shorter tracking latencies when the original stimulus array

375   was present compared to when it was absent, but we observed no difference in tracking accuracy

376   or latency between the Self and Other conditions. As we note above, one possibility is that the

377   Self versus Other manipulation may have been too subtle. To address this uncertainty, we tested a

378   more extreme tracking manipulation. In Experiment 2, participants completed the same task,

379   except instead of viewing another participant's eye movements in the Other condition, we

380   presented their own temporally reversed eye movements during the second task. In other words,

381   we replaced the Other condition with a Reverse scanpath condition (Figure 1F). This way, we

382   can control for differences in variables like fixation duration, saccade amplitude, and velocity

383   between the conditions. An exploratory analysis of the data from Experiment 1 (Supplemental

384   Figure 2) found individual differences in saccade amplitude to be a mediating factor for tracking

385   accuracy ($r(22) = -0.44$, $p = 0.03$), with reduced tracking accuracy when the median saccade

386   amplitude in the replayed data were large.

387   In addition, as each display in Experiment 1 was randomly generated, we did not have a

388   way to directly assess consistency in gaze behaviour within each participant (i.e., whether

389   participants tend to look in the same locations and in a similar order). We modified this in

390   Experiment 2, in which each search display consisted of one of ten unique spatial arrangements

391   (i.e., layouts), with targets and distractors assigned to random locations within each layout

392   (Figure 1E). This allowed us to assess consistency by comparing scanning behaviour for displays

393   in which items were shown in the same spatial configuration. In an exploratory analysis, we

394   tested whether participants who were more consistent in their scanning behaviour would be more

395   affected by the tracking manipulation (i.e., Forward versus Reverse).

396   **Methods**

397   **Participants.**

398   Twenty-nine participants were recruited from the participant pool at the University of

399   Toronto Mississauga. Following the same exclusions in Experiment 1 (see *Exclusion Criteria*),

400   the final sample consisted of twenty-four participants (age: $M = 20$, 6 males, 17 females, and 1

401   unreported). Four participants were excluded due to a high proportion of lost eye-tracking

402   samples (> 20%), and one participant did not finish the study and their data were discarded. Like

403   Experiment 1, all participants provided informed consent, as approved by the Research Ethics

404   Board at the University of Toronto and were compensated at $15 CAD/hour (average study time:

405   75 to 90 minutes). Eligibility requirements are also similar to Experiment 1, with the addition

406   that we excluded any participant from Experiment 1 to avoid overfamiliarity with the task and to

407   preserve participant naivety.

408   **Materials and Stimuli.**

409   Apart from the following changes, Experiment 2 followed an identical procedure to

410   Experiment 1.

411   *Spatial arrangements.* To investigate the consistency of gaze behaviour within

412   individuals, ten spatial arrangements were randomly generated prior to data collection (Figure

413   1E). This was done to control for the items' locations (i.e., the locations of the "slots" that items

414   could be in within an invisible grid), but to vary *what* was placed at each location. Each spatial

415   arrangement consisted of 25 allocated positions in an invisible 8×11 grid, with random

416   horizontal and vertical spatial jitters added. Unlike in Experiment 1, where we presented 180

417   uniquely random stimulus arrays to each participant, such that no participant saw the exact same

418   set of stimulus arrays, we presented the same ten spatial arrangements to all participants multiple

419   times. Each iteration was randomly selected on each trial, which resulted in approximately 16-20

420   iterations per spatial arrangement since there were 180 trials in total, including practice trials.

421   Importantly, only the item "slots" remained constant across each repetition. Each display had a

422   different set of items (target and distractors) assigned to these positions, with varying

423   orientations and greyscale intensity values. In this manner, on target-present trials (50% of the

424   trials), the target was in a random location, and only the spatial arrangement of the stimuli was

425   held constant. Aside from the spatial locations, each iteration (i.e., stimulus array) was random

426   and unique from the other.

427   *Tracking conditions.* As in Experiment 1, participants completed a tracking task in which

428   they followed a moving red dot on the screen with their eyes. However, instead of viewing either

429   their own eye movements or that of the previous participant, participants in Experiment 2 only

430   viewed their own. On each trial, the dot replayed the original, unaltered sequence of positions

431   (Forward condition) or a temporally reversed one (Reverse condition), with the same 50%

432   visibility manipulation used in Experiment 1 (see *Materials and Stimuli*). This allowed us to

433   match individual gaze metrics (i.e., saccade amplitude, fixation duration) between these

434   conditions and minimize noise due to between-subject differences. As such, excluded

435   participants were no longer replaced in a staggered manner. Apart from this difference, the task

436   proceeded identically to Experiment 1. As shown in Figure 1F, we used a 2×2 design, with the

437   factors of tracking condition (Forward versus Reverse) and stimulus array (Background versus

438   No Background). As in Experiment 1, trials were randomly interleaved and participants were not

439   informed of this manipulation, or that the tracking task would follow the search task.

440   **Analysis.**

441   *Consistency analysis.* In addition to our main analysis described in Experiment 1 (*see*

442   *Analysis*), we employed a secondary analysis based on the ten spatial configurations in our

443   experimental design. To extract a measure of participants' consistency in scanning the displays,

444   gaze positions over time for each spatial configuration were correlated across all possible pairs of

445   trials with that spatial configuration. This was done separately for the x-coordinates and y-

446   coordinates, and then each set of correlation values underwent Fisher Z transformation prior to

447   taking the average of both x- and y-gaze position correlations. This resulted in ten average

448   correlation values, one for each layout. To estimate overall within-subject consistency, we took

449   the average of all ten correlation values. Higher correlation values indicated higher consistency

450   (i.e., participants reliably looked at the same locations in each iteration of the spatial

451   configuration in a similar sequence).

452   **Results**

453   **Tracking accuracy.** Using the same analysis in Experiment 1, a 2 (condition: Forward

454   versus Reverse) × 2 (stimulus array: Background versus No Background) repeated-measures

455   ANOVA revealed a significant difference between the Forward and Reverse conditions on

456   tracking accuracy, $F(1,23) = 16.57$, $p < .001$, $\eta^2_p = 0.42$. As shown in Figure 4A, participants

457   were more accurate at tracking an unaltered sequence of their own eye movements (Forward; M

458    $= 0.97$, $SD = 0.10$) compared to when the sequence was temporally reversed (Reverse; $M = 0.92$,

459    $SD = 0.09$). Similarly, we observed a main effect of stimulus array presence on tracking

460   accuracy, $F(1,23) = 8.55$, $p = .008$, $\eta^2_p = 0.27$, wherein the presence of the original stimulus array

461   facilitated more accurate tracking (Background; $M = 0.96$, $SD = 0.10$) compared to when it was

462   absent (No Background; $M = 0.93$, $SD = 0.09$). However, there was no interaction between these

463   factors, $F(1,23) = 0.18$, $p = .678$.

464   **Tracking latency.** We found similar results to Experiment 1, as shown in Figure 4B. On

465   average, participants had lower tracking latencies when the background was present

466   (Background; $M = 181.92$ ms, $SD = 36.12$ ms) versus when it was absent (No Background; $M =$

467   206.76 ms, $SD = 38.35$ ms), $F(1,23) = 23.57$, $p < .001$, $\eta^2_p = 0.50$. However, the tracking

468   manipulation (Forward versus Reverse) did not affect latency, $F(1,23) = 1.53$, $p = .23$, ($M =$

469   191.14, ms SD = 40.48 ms versus M = 197.55 ms, SD = 37.84 ms). There was also no significant

470   interaction, $F(1,23) = 0.18$, $p = .678$. In sum, these results suggest that extreme manipulations

471   (i.e., the Forward versus Reverse conditions) impacted accuracy, but not latency.

472   **Relationship between consistency and tracking performance.** In an exploratory

473   analysis, we determined whether consistency in scanning behaviour modulated the difference in

474   tracking performance between the Forward and Reverse conditions. To do this, we first

475   calculated the within-subject consistency (see Experiment 2, *Correlation Analysis*). Higher

476   positive Fisher Z-transformed correlation values indicate greater consistency, while values closer

477   to zero indicate more random gaze behaviour. The correlation analysis revealed a range of

478   individual differences in consistency (Figure 5A); across participants, correlation values ranged

479   from 0.02 to 0.54.

480   Similarly, with regards to tracking performance, some participants had a larger reduction

481   in tracking performance in the Reverse condition compared to the Forward manipulation than

482   others. We investigated whether consistent individuals would have a larger decrement in tracking

483   performance with a temporally reversed compared to unaltered scan path. We represented this

484   susceptibility as a *Forward – Reverse* difference in tracking accuracy and latency, then correlated

485   these values with their corresponding within-subject consistency values using a Pearson

486   correlation. We found that within-subject consistency was associated with this difference in

487   tracking latency, $r(22) = -.60$, $p = .002$ (Figure 5C) such that more consistent participants

488   showed a larger decrement in tracking latency for Forward compared to Reverse scan paths (i.e.,

489   they were faster in the Forward compared to the Reverse condition). However, there was no

490   relationship between consistency and the difference in accuracy between the Forward and

491   Reverse conditions, $r(22) = .17$, $p = .42$ (Figure 5B). In sum, our correlation analysis suggests

492   that people who are more internally consistent in their scanning behavior are more affected by

493   the temporal reversal of their eye movement sequence, as evident by the latency benefit

494   (negative latency values on the x-axis, Figure 5C), while those who are less consistent seemed to

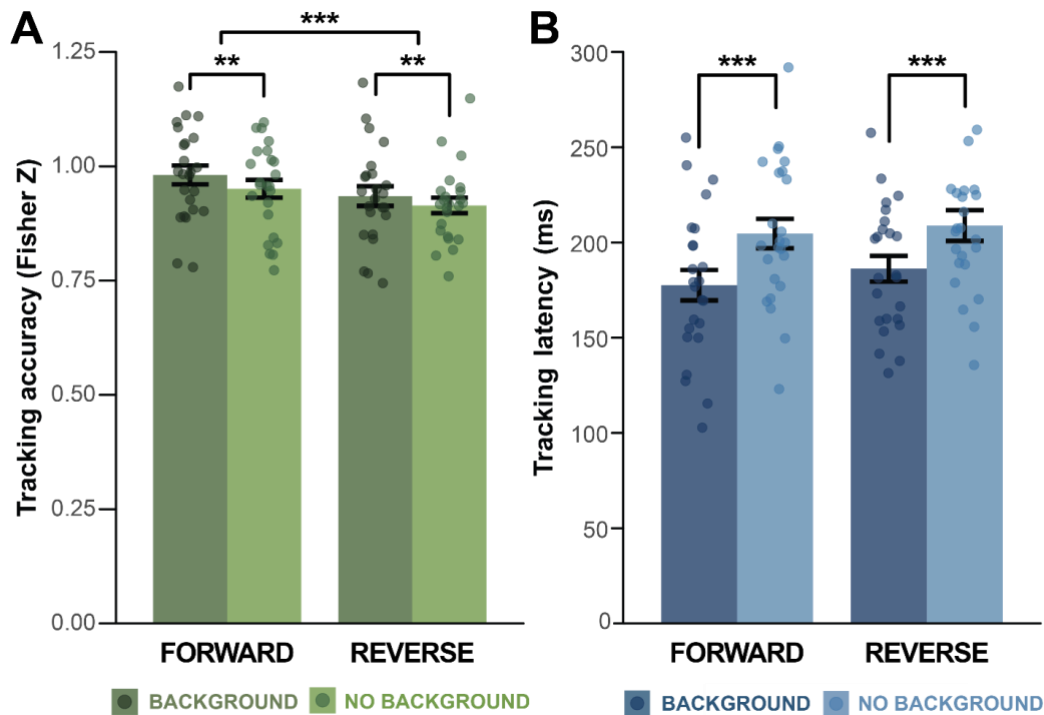495   be less affected.

496
497   **Figure 4.** Tracking performance in Experiment 2. (A) Average tracking accuracy in each condition,
498   calculated as the peak of the cross-correlation (Fisher Z-transformed correlation value). (B) Average tracking
499   latency in each condition, calculated as the lag corresponding to the peak accuracy (i.e., the point on the x-
500   axis of the CCG corresponding to the peak). Symbols represent individual participants, and error bars
501   represent $\pm$ 1 standard error of the mean (SEM). Asterisks represent results of the 2×2 repeated-measures
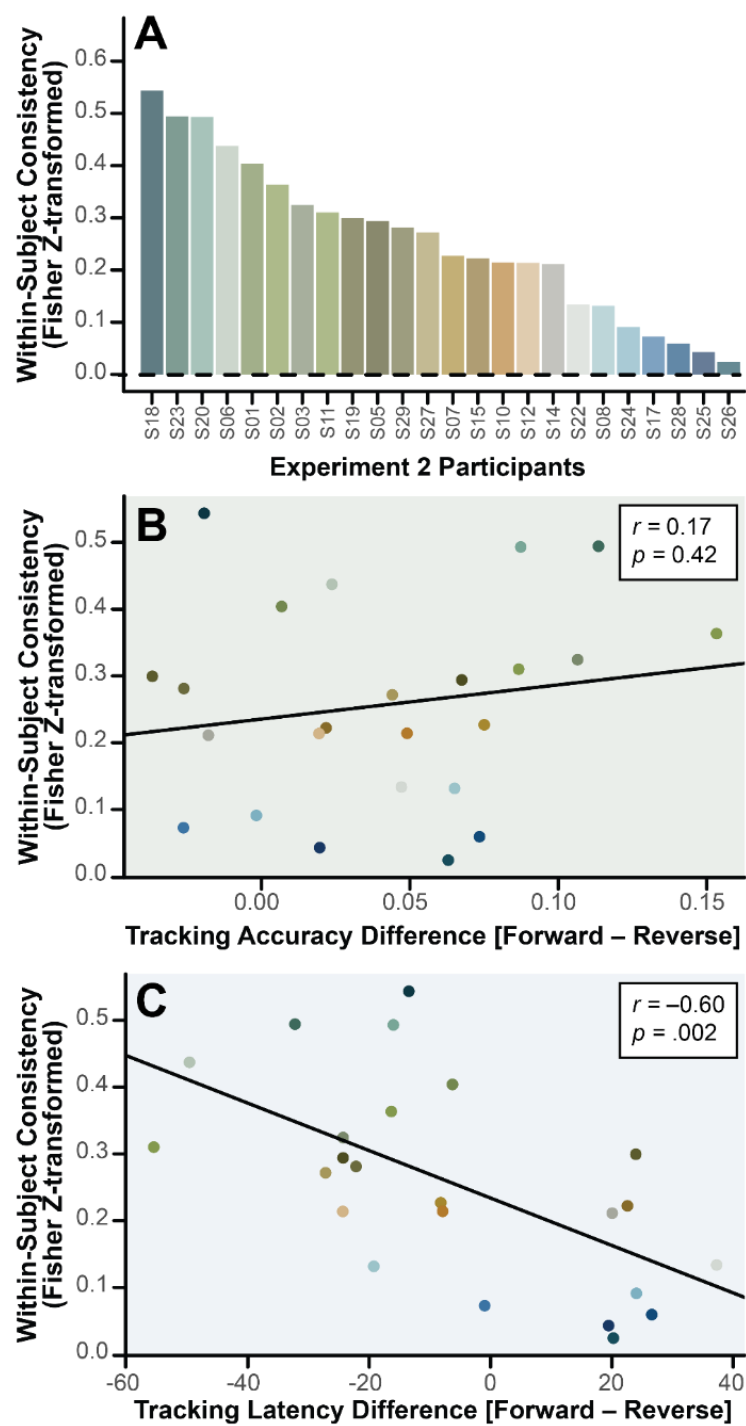502   ANOVA, ***$p < .001$; **$p < .01$.

504 **Figure 5. Individual variability in within-subject consistency. (A)** Within-subject consistency
505 for each participant in Experiment 2. Each bar represents the mean within-subject consistency,
506 calculated from the average of all ten spatial arrangements (Fisher Z-transformed correlation
507 value). **(B)** Correlation between *Forward – Reverse* difference in tracking accuracy (x-axis) for
508 all participants with their within-subject consistency (y-axis). Negative x-axis values mean that the
509 participant performed better during the Reverse conditions, while positive values mean that they
510 performed better during the Forward conditions. Scatter points represent individual participants.
511 **(C)** Correlation comparing the *Forward – Reverse* difference in tracking latency (x-axis) with their
512 within-subject consistency (y-axis). Negative x-axis values correspond to shorter latencies (i.e.,
513 faster tracking) in the Forward compared to the Reverse condition, while positive values
514 correspond to shorter latencies in the Reverse compared to Forward condition. Scatter points
515 represent individual participants.

Discussion

517   Experiment 2 demonstrated that temporally reversing the eye movement sequence

518 produced decrements in tracking accuracy, but did not affect latency. This could be explained by

519 the fact that this manipulation flipped the direction of the eye movement sequence. For example,

520 participants may retain some information about the direction in which they tend to scan the

521 display (e.g., clockwise) and rely on this information during the brief segments when the

522 replayed dot disappears, which would drive down the tracking accuracy measure.

523   We also examined whether people with more consistent scanning behaviour would be

524 more affected by the Forward versus Reverse manipulation. We operationalized this

525 susceptibility to the temporal reversal as a *Forward – Reverse* difference in accuracy and latency.

526 We found a negative correlation between susceptibility (in our latency measure) and within-

527 subject consistency, such that people who were more consistent had greater latency decrements

528 in the Forward compared to the Reverse conditions. However, we did not see a similar

529 relationship with tracking accuracy (i.e., within-subject consistency was not associated with the

530 effect of the manipulation on tracking accuracy). Regardless, we demonstrated that with a

531 sufficiently extreme change in the path of the tracked dot, participants' tracking performance

532   decreased. This could indicate that the null result we observed in Experiment 1 may a

533   consequence of using a more subtle manipulation (Self versus Other).

534   **Experiment 3**

535   Experiment 2 demonstrated that a highly dissimilar eye movement sequence (i.e.,

536   temporal reversal) is more difficult for participants to track than their own gaze sequence. These

537   results are consistent with the idea that we have some information about our own typical gaze

538   behaviour. This addresses one of the possible interpretations we had for the results in Experiment

539   1, supporting the idea that the Self versus Other manipulation in Experiment 1 may not have

540   been large enough to produce differences in tracking performance. For example, it is possible

541   that we observed no difference in tracking performance in Experiment 1 because highly similar

542   people were paired by random chance. What remains uncertain is whether tracking performance

543   could, in principle, be sensitive to individual differences in gaze behaviour.

544   To resolve this question, in Experiment 3, we had three conditions. In one, participants

545   tracked their own previously recorded eye movements. We also had two different "other"

546   conditions: one that replayed eye movements from a participant who was very similar to the

547   current participant, and another that replayed eye movements from a participant who was very

548   dissimilar. The tracked positions for these two "other" conditions ("Other-Most Similar" and

549   "Other-Least Similar", respectively) were selected from participants who had previously

550   completed Experiment 2.

551   **Methods**

552   **Participants**

553   Thirty-two participants were recruited from the paid participant pool at the University of

554   Toronto Mississauga. Eligibility requirements and inclusion criteria were held constant from

555   Experiments 1 and 2, with the addition that we excluded any participant who had done either of

556   the two previous experiments to preserve task naivety. Following exclusions (see Experiment 1,

557   *Exclusion Criteria*), the final sample consisted of twenty-four participants (age: $M = 20$, 6 males,

558   18 females). Six participants were excluded due to a high proportion of lost eye-tracking samples

559   (> 20%), one was excluded due to low accuracy (not significantly above chance based on a

560   binomial test), and another was excluded due to a low median reaction time (less than 3.5

561   seconds). Four participants were compensated with course credit in accordance with their

562   request, while the rest were compensated at $15 CAD/hour. In general, participants completed

563   the experiment in 75 to 90 minutes.

564   **Materials and Stimuli**

565   Apart from the following changes, Experiment 3 followed an identical procedure to

566   Experiment 2, including the same set of ten spatial arrangements.

567   *Tracking conditions*. To isolate the effects of scanpath similarity within our original

568   design, we measured participants' tracking performance for their own gaze positions ("Self"

569   condition) and compared this with two different "other" tracking conditions: "Other-Most

570   Similar" and "Other-Least Similar". These were taken from the previously recorded gaze data

571   from participants who completed Experiment 2, since the same set of spatial arrangements were

572   used in both experiments. As shown in Figure 6B, after each participant completed the search

573   task, but prior to starting the tracking task, we correlated the x- and y- gaze positions between the

574   current participant (in Experiment 3) and each participant from the previous set of 24 participants

575   (Experiment 2). Pairwise correlations between the current participant and the Experiment 2

576   participants were first calculated separately for each matched spatial arrangement (i.e., layout),

577   and then averaged across layouts to produce an overall measure of similarity between the current

578   participant and each Experiment 2 participant.

579   Next, these resulting correlation values were used to determine the participants from

580   Experiment 2 to use for the Other-Most and Other-Least replay conditions in Experiment 3. To

581   minimize the odds of always selecting the same participant for the "Other-Most" and "Other-

582   Least" conditions, we randomly selected one out of four of the most correlated participants for

583   the Other-Most condition, and one out of the four least correlated participants for the Other-Least

584   condition. In addition to these two "other" conditions, we included the original Self condition,

585   wherein we replayed the current participant's own eye movements. Concisely, in this version, we

586   tested three conditions in the second half of the experiment: Self, Other-Most, and Other-Least.

587   Unlike Experiments 1 and 2, every trial showed a replay of the gaze positions superimposed on

588   the original stimulus array.

589   As shown in Figure 6A, in the first half of the experiment, participants completed 30

590   practice trials and 150 experimental trials of the search task. In the second half (Tracking task),

591   participants completed 30 practice trials followed by 150 tracking trials, which consisted of 50

592   experimental trials per condition, which were randomly interleaved. As in the previous

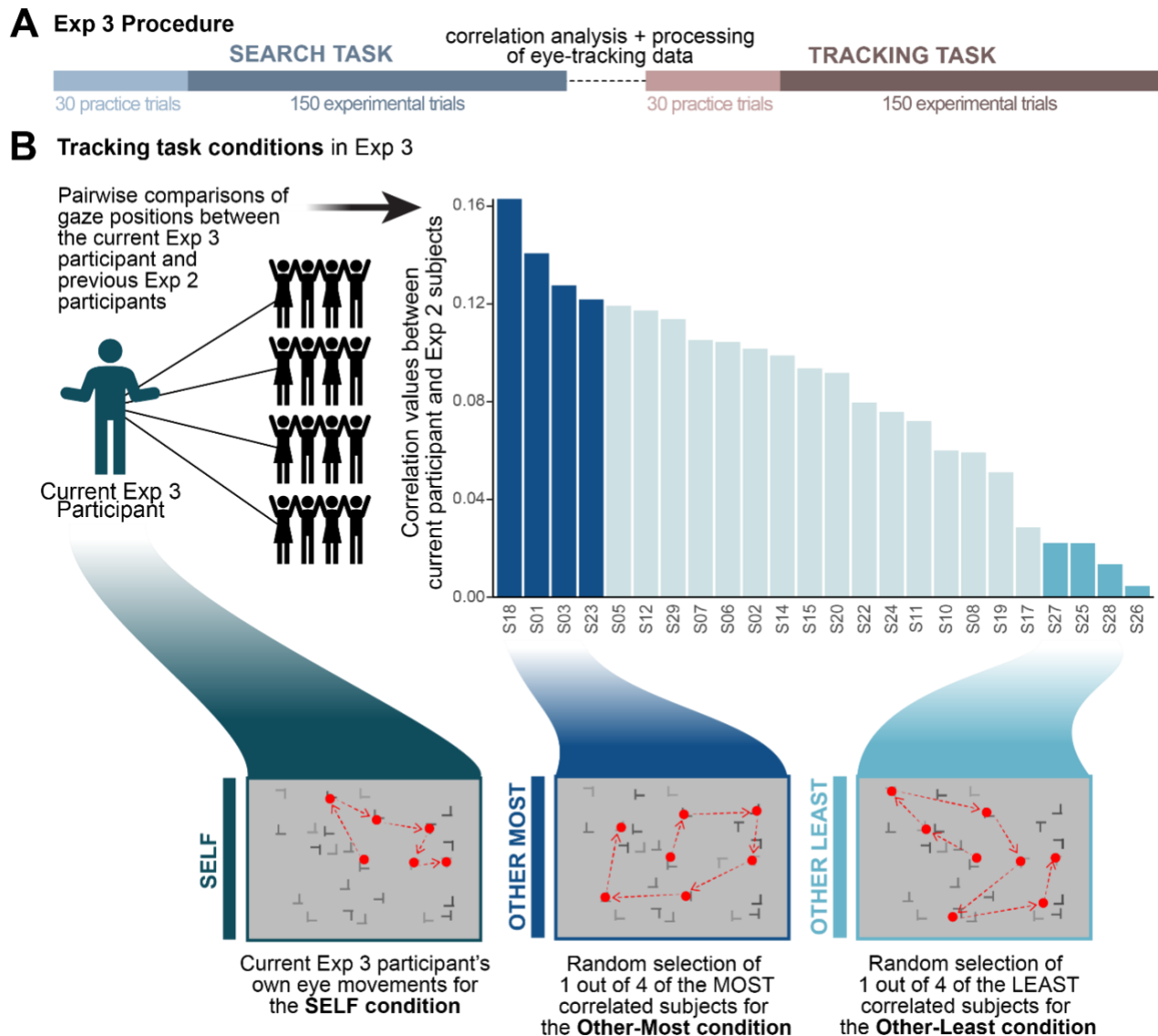593   experiments, participants were not informed of the different tracking conditions.

A Exp 3 Procedure

B Tracking task conditions in Exp 3

594
595 **Figure 6. Design and procedure for Experiment 3.** (A) *Schematic of trial order*. Participants
596 completed 30 practice trials followed by 150 trials of the experiment proper for both tasks, with an option
597 for breaks after every 40 trials. Between the two tasks, a longer break occurred in which we processed the
598 eye-tracking data and prepared it for the tracking task. (B) *Conditions for the tracking task, Experiment 3*.
599 The determination of the two "Other" conditions was based on pairwise comparisons between the current
600 Experiment 3 participant's gaze data and previous participants from Experiment 2. Each bar represents the
601 calculated Fisher Z-transformed correlation value between the current participant and each Experiment 2
602 participant (e.g., S18 or Subject 18, was the most correlated for this current participant). The dark blue bars
603 represent the four most correlated Experiment 2 participants, and the light blue bars represent the four least
604 correlated. From each of these, one was randomly selected for the Other-Most Similar ("Other-Most") and
605 Other-Least Similar ("Other-Least) conditions, respectively (e.g., S23 for Other-Most and S25 for Other-
606 Least). In a random order, participants were presented with the dot replaying eye movements from
607 themselves (Self), from someone who was very similar to them (Other-Most), or from someone who was
608 very dissimilar to them (Other-Least). The original stimulus array was present in all trials for Experiment 3.

609   **Results**

610   **Tracking accuracy.** We repeated the same analyses from Experiments 1 and 2. A one-

611   way repeated-measures ANOVA found no main effect of eye movement source on tracking

612   accuracy, $F(2,46) = 0.82$, $p = .446$. As shown in Figure 7A, on average, participants were no

613   more accurate at tracking the red dot, regardless of whether it replayed their own eye movements

614   (Self; $M = 0.95$, $SD= 0.12$), someone with similar scanning behaviour (Other-Most; $M = 0.97$,

615    $SD= 0.11$), or someone else with very dissimilar scanning behaviour (Other-Least; $M = 0.96$, SD

616    $= 0.10$).

617   **Tracking latency.** With respect to tracking latency (Figure 7B), a separate one-way

618   repeated-measures ANOVA found a main effect of eye movement source, $F(2,46) = 6.88$, $p =$

619    $.002$, $\eta^2_p = 0.230$. Post-hoc Tukey's tests further revealed that, on average, participants had a

620   lower latency when tracking replayed Other-Most eye movements compared to when they were

621   tracking their own eye movements (Self), $t(23) = -3.07$, $p_{Tukey} = .014$. Similarly, participants had

622   shorter tracking latencies for the Other-Most condition compared to the Other-Least condition,

623    $t(23) = 3.83$, $p_{Tukey} = .002$. However, no difference in latency was found between Self and Other-

624   Least conditions, $t(23) = 0.51$, $p_{Tukey} = .866$.

625   In sum, surprisingly, participants had a shorter tracking latency when tracking the eye

626   movements of the participant from Exp 2 most similar to themselves (Other-Most; $M = 164.74$

627   ms, $SD= 35.60$ ms), compared to either themselves (Self; $M= 189.57$ ms, $SD= 32.70$ ms) or the

628   participant they were least similar to (Other-Least; $M = 194.62$ ms, $SD = 43.96$ ms).

629   **Relationship between consistency and tracking performance.** In an exploratory

630   analysis, we repeated the analyses from Experiment 2 (see *Correlation Analysis* and *Results*) and

631   calculated a within-subject consistency measure for each participant. In Figure 8A, we observed

632   a large range of differences in internal consistency across participants, which led us to examine

633   whether this was related to the estimates of similarity and dissimilarity when assigning

634   conditions (i.e., Other-Most and Other-Least). One possibility is that, for a participant with

635   relatively random and less consistent gaze behaviour, that they might only be highly correlated

636   with another participant with a relatively high within-subject consistency measure. In other

637   words, the participant selected for the "Other-Most" condition would not only be similar to them,

638   but also have a high degree of internal consistency. We asked whether this would explain the

639   counterintuitive results we found in Experiment 3, specifically, the observation that participants

640   had shorter latencies for the Other-Most condition compared to the Self condition. We postulated

641   that instead of gaze behaviour similarity, consistency in scanning behaviour modulated these

642   results.

643   To assess this, we first calculated each participant's consistency in Experiment 3,

644   following the same analyses outlined for Experiment 2 (Figure 8A). Next, we calculated the

645   between-subject consistency difference (*Other-Most – Self* difference in within-subject

646   consistency). We then calculated the observed latency difference between the Self and Other-

647   Most conditions in Experiment 3, and then correlated it with the corresponding between-subject

648   consistency difference using a Pearson correlation (Figure 8B). We found a negative correlation

649   between the observed latency difference and the between-subject consistency difference, $r(22) =$

650     $-.59, p = .003$. This would indicate that tracking a consistent scanpath, regardless of the

651   condition, resulted in shorter latencies. More precisely, when the current participant (Self) was

652   less consistent than the Other-Most subject, tracking latency was lower for the Other-Most

653   condition; and vice versa. To summarize, our correlation analysis suggests that people who are

654   more internally consistent may be easier for another participant to track, as evident by the shorter

655   latencies (negative latency values on the x-axis, Figure 8C), while those who are more random in

656   their gaze behaviour may be harder for another participant to track. However, we did not observe

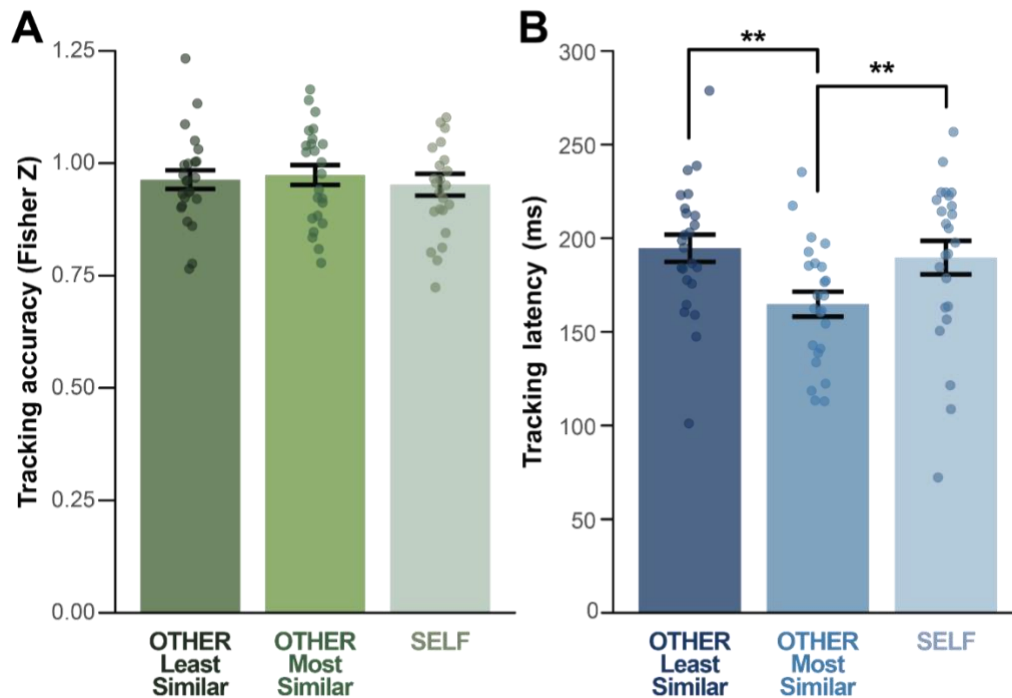657   this relationship for tracking accuracy, $r(22) = 0.90$, $p = .68$ (Figure 8B).

658

659   **Figure 7.** Tracking performance in Experiment 3. (A) Average tracking accuracy in each condition, calculated
660   as the peak of the cross-correlation (Fisher Z-transformed correlation value). (B) Average tracking latency in
661   each condition, calculated as the lag corresponding to the peak accuracy (i.e., the point on the x-axis of the
662   CCG corresponding to the peak). Symbols represent individual participants, and error bars represent ± 1
663   standard error of the mean (SEM). Asterisks represent results of the one-way ANOVA, **$p < .01$.
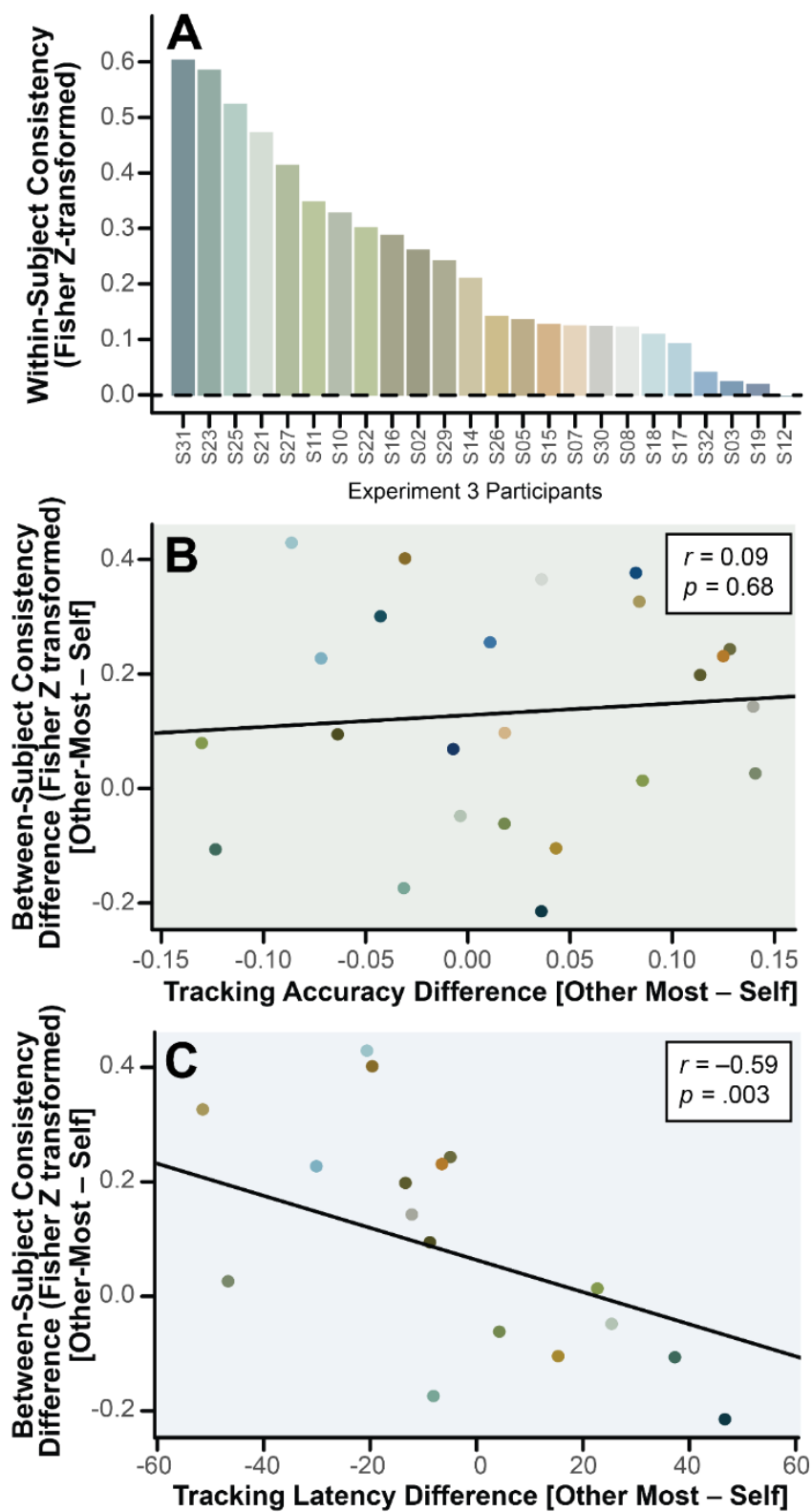
Figure 8. Within-subject consistency modulates the tracking latency difference between Other-Most and Self. (A) Within-subject consistency for each participant in Experiment 3. Each bar represents the mean within-subject consistency, calculated from the average of all ten spatial arrangements (Fisher Z-transformed correlation value). Correlation values ranged from – 0.0012 to 0.61 across all participants. **(B)** Correlation between the difference in internal consistency between the Other-Most and Self conditions (y-axis), and the tracking accuracy difference between these two conditions (x-axis). Negative values on the x-axis indicate that the participant had higher tracking accuracy in the Self condition compared to the Other-Most Condition, while positive values correspond to higher accuracy in the Other-Most compared to the Self condition. On the y-axis, negative values mean that the current participant had higher internal consistency, while positive values indicate that the Other-Most participant was more consistent internally. Scatter points represent individual participants. **(C)** Correlation between the difference in internal consistency between the Other-Most and Self conditions (y-axis), and the tracking latency difference between these two conditions (x-axis), following the same conventions as panel B. Negative values on the x-axis mean that the participant had shorter latencies (faster tracking) in the Other-Most compared to the Self condition, while positive values correspond to shorter latencies in the Self compared to the Other-Most condition.

Discussion

Experiment 3 showed no difference in tracking accuracy between the three conditions: Self, Other-Most, and Other-Least. However, we did find a significant latency advantage for Other-Most trials compared to the other two conditions. Intuitively, we expected that Other-Most and Self would have similar accuracy or latency advantages, based on the results of Experiment 2, which suggested that participants' tracking performance might be worse for scanpaths that are very different from their own. Surprisingly, our results showed that participants had significantly shorter latencies in the Other-Most condition compared to the Self condition. As shown by our correlation analysis (Figure 8B), one factor that may account for this difference between Other-Most and Self may be a difference in the within-subject consistency.

Together, this would suggest that the difference in latency may be an artifact of the selection process for each condition, which would produce differences in internal consistency as well as between-subject similarity between conditions. By choosing Experiment 2 participants for the "Other-Most condition" based on their correlation with the current participant, this

696   process may have unintentionally also selected highly consistent participants for the "Other-

697   Most" condition relative to the other two conditions. A post-hoc *t*-test revealed significantly

698   higher consistency values for the Other-Most compared to the Self condition, further supporting

699   this possibility, $t(23) = 3.42$, $p = .002$. Similarly, the lack of latency difference between the Self

700   and Other-Least conditions may be due to the randomness of the Self and Other-Least

701   participant's gaze behaviour. That is, the Other-Least participants could be weakly correlated

702   with the current participant either because their gaze behaviour was relatively random, or

703   because their behaviour was consistent, but different from the current participant. We note that

704   internal consistency was somewhat higher for the Self condition compared to the Other-Least

705   conditions $t(23) = -2.31$, $p = .03$, supporting the former interpretation. In sum, Experiment 3

706   revealed that internal consistency modulated tracking latency differences, which also aligns with

707   the results we observed in Experiment 2 (Figure 5C). As the design of Experiment 3 did not

708   allow us to fully separate similarity from consistency, we cannot determine the extent to which

709   eye movement similarity underlies differences in tracking accuracy between conditions.

General Discussion

711   In this study, we investigated whether participants retain some information about their

712   typical gaze behaviour using a novel tracking paradigm as an indirect probe. To briefly

713   summarize the results of Experiment 1, we investigated whether participants would have better

714   tracking performance (i.e., higher accuracy or shorter latency) when tracking their own eye

715   movements compared to those of another person. We found no differences in tracking

716   performance when participants tracked their own eye movements compared to those of another

717   participant (Self versus Other conditions, respectively). However, we found a latency benefit

718   with the presence of the original stimulus array. In Experiment 2, we demonstrated that reversing

719   the temporal sequence of eye movements was sufficient to producing a tracking accuracy

720   decrement, such that participants were less accurate at tracking temporally reversed scan paths

721   compared to unaltered ones (Reverse versus Forward, respectively). Simultaneously, we

722   observed a persistent latency benefit as well as a new accuracy benefit with the presence of the

723   original stimulus array, which was largely consistent with our results from Experiment 1. This

724   prompted us to revisit the Self versus Other comparison in Experiment 3, in which participants

725   tracked a dot that replayed gaze positions in one of three conditions: participants' own previously

726   recorded eye movements (Self), those of a participant highly correlated with themselves (Other-

727   Most), or those of a participant poorly correlated with themselves (Other-Least). In Experiment

728   3, while we continued to find no differences in tracking accuracy across conditions, a surprising

729   latency advantage was revealed in the Other-Most conditions compared to Self and Other-Least.

730   This result could reflect differences in internal consistency, rather than similarity per se.

731   Our results, specifically the Self versus Other comparisons in Experiments 1 and 3,

732   seemed contrary to previous literature on stable, large individual differences in gaze behaviour

733   within visual search tasks. Gaze behaviour differences between individuals can be attributed to

734   an amalgamation of high-level influences, like strategy (Boot et al., 2009; Clarke et al., n.d., 2022;

735   Gilchrist & Harvey, 2006), and low-level factors, like individual visual sensitivity (Veríssimo et al.,

736   2021). These different factors have been known to impact the temporal order, spatial distribution,

737   and number the eye movements we make. One possible explanation for the lack of tracking

738   performance differences between Self and Other is due to having relatively poor knowledge of

739   their own eye movements, which would be consistent with previous literature (Clarke et al., 2017;

740   Foulsham & Kingstone, 2013; Kok et al., 2017; Marti et al., 2015; Van Wermeskerken et al., 2018; Võ et

741   al., 2016). Another possibility is that the task might impact the degree to which individuals exhibit

742   differences in gaze behaviour. Some tasks could drive down consistent and reliable gaze

743   behaviour differences between individuals, due to the visual information in the environment or

744   the nature of the task, which would prevent participants from tracking their own eye movements

745   better or faster those of another participant. Our results in Experiment 2 would be consistent with

746   the latter assertion, such that when we induced extreme changes to the replayed eye movements

747   (i.e., Forward versus Reverse), we were able to induce tracking performance differences.

748   Nevertheless, the threshold of our limited awareness remains unclear. How extreme must

749   these differences be for us to reliably detect them? Further work using different stimuli to induce

750   larger idiosyncratic behaviours compared to a search for Ts among Ls, may be required.

751   Regardless, at least in the context of the task tested here, the observed tracking differences in our

752   results suggest that our awareness of our own gaze behaviour is relatively limited when probed

753   with an implicit task.

754   A further consideration is whether this method captures the extent to which participants

755   retain information (either implicitly or explicitly) about their own gaze behavior. For example,

756   one possibility is that participants could achieve a high level of tracking accuracy, not by

757   tracking the dot, but instead by simply be reproducing their own previous gaze behaviour (as

758   might be observed if the dot were absent for the duration of the entire trial). However, we note

759   that the latencies we observed are consistent with latencies of reflexive saccades to a target (on

760   the order of 200 ms), suggesting that participants are indeed tracking the position of the dot.

761   Another interpretation of the result we observed in Experiment 2 is that participants might track

762   the dot more accurately in the Forward compared to the Reverse condition due to differences in

763   saccadic precision for different saccade directions. In other words, participants might naturally

764   adopt particular saccade directions (i.e., in the Forward condition) because those directions tend

765   to produce lower errors to begin with. This interpretation is unlikely, because scanning the search

766   display requires making saccades in multiple directions, and finer spatial precision for one

767   direction would be counteracted by lower precision in another. A final possibility is that

768   differences between the Forward and Reverse conditions may reflect differences in search habits

769   rather than participants' knowledge their own gaze behaviour. For example, participants may

770   revert back to their typical scanning behaviour in the Reverse condition whenever the dot

771   disappears, which would reduce tracking accuracy. Nonetheless, this interpretation would be

772   consistent with the idea that participants retain some information, albeit implicitly, about typical

773   scanning behaviour.

774   Together, the results of these experiments also indicate that a number of factors may

775   influence performance in this tracking task, making it challenging to use as a method for probing

776   awareness. In particular, both Experiments 2 and 3 support the idea that tracking latency

777   decreases when participants are required to track a consistent scan path. In addition, Experiment

778   3 showed that in the span of thirty minutes to an hour (average time for the second task),

779   participants were able to extract information about another participant's consistent gaze

780   behaviour which may be a product of statistical learning. Statistical learning is learning based on

781   the extraction of environmental patterns, regularities, and distribution (Perruchet & Pacton, 2006;

782   Sznabel et al., 2023), and may explain why participants were able to reproduce and track the eye

783   movements of other people with shorter latencies when those individuals were highly consistent

784   (Figure 5C, 8C). Having more consistent scanning behaviour produced reliable patterns for

785   individuals to extract and learn implicitly. This type of learning would be relevant to the growing

786   interest in using gaze behaviour and modelling to guide trainees in radiology, airport baggage

787   screening, and other real-world search tasks.

788   A further area for future studies is to examine how scanning behaviour would interact

789   with search performance and target detection. In our study, we showed that even though

790   participants' knowledge of their own gaze behaviour was limited, people were able to learn

791   others' gaze behaviours and track them with a shorter delay. Future work might examine the

792   question of consistency in more detail, specifically in the context of search performance,

793   focusing on whether having a consistent scanpath would be more helpful or harmful to overall

794   search performance.

795   **Declarations**

796   **Availability of Data and Materials**

797   All data and materials are available on OSF [https://osf.io/gw4uf/]. All experiments were

798   preregistered on OSF at [https://osf.io/gw4uf/].

799   **Code availability**

800   Code for stimulus presentation and analysis is available on OSF at [https://osf.io/gw4uf/].

801   **Funding**

802   This work was supported by an undergraduate research grant from the University of Toronto

803   Mississauga to AHC, and an Natural Sciences and Engineering Research Council (NSERC) Discovery

804   Grant to Anna Kosovicheva (RGPIN-2022-03131).

805   **Conflicts of Interest**

806   The authors have no conflicts of interest to declare.

807   **Ethics Approval**

808   This research was approved by the Research Ethics Board at the University of Toronto (#41533).

809   **Consent to participate**

810   All participants provided written informed consent prior to participating in the study.

811   **Consent for publication**

812   Participants consented to the publication of their deidentified data.

813   **Author contributions**

814   AHC and AK conceived and programmed the experiments. AHC collected the data. AHC and AK analyzed

815   the data, interpreted the results, and wrote the manuscript.

References

816

817   Abrams, R. A., & Jonides, J. (1988). Programming saccadic eye movements. *Journal of Experimental*

818   *Psychology: Human Perception and Performance*, *14*(3), 428–443. https://doi.org/10.1037/0096-

819   1523.14.3.428

820   Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when

821   viewing different visual environments. *Vision Research*, *39*(17), 2947–2953.

822   https://doi.org/10.1016/S0042-6989(99)00019-X

823   Ashraf, H., Sodergren, M. H., Merali, N., Mylonas, G., Singh, H., & Darzi, A. (2018). Eye-tracking

824   technology in medical education: A systematic review. *Medical Teacher*, *40*(1), 62–69.

825   https://doi.org/10.1080/0142159X.2017.1391373

826   Boot, W. R., Becic, E., & Kramer, A. F. (2009). Stable individual differences in search strategy?: The effect

827   of task demands and motivational factors on scanning strategy in visual search. *Journal of Vision*,

828   *9*(3), 1–16.

829   Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

830   Carpenter, R. H. S. (1988). *Movements of the eyes, 2nd rev. & enlarged ed* (p. 593). Pion Limited.

831   Clarke, A. D. F., Irons, J. L., James, W., Leber, A. B., & Hunt, A. R. (n.d.). Stable individual differences in

832   strategies within, but not between, visual search tasks. *Quarterly Journal of Experimental*

833   *Psychology*.

834   Clarke, A. D. F., Mahon, A., Irvine, A., & Hunt, A. R. (2017). People Are Unable to Recognize or Report on

835   Their Own Eye Movements. *Quarterly Journal of Experimental Psychology*, *70*(11), 2251–2270.

836   https://doi.org/10.1080/17470218.2016.1231208

837   Clarke, A. D. F., Nowakowska, A., & Hunt, A. R. (2022). Visual search habits and the spatial structure of

838   scenes. *Attention, Perception, & Psychophysics*, *84*(6), 1874–1885. https://doi.org/10.3758/s13414-
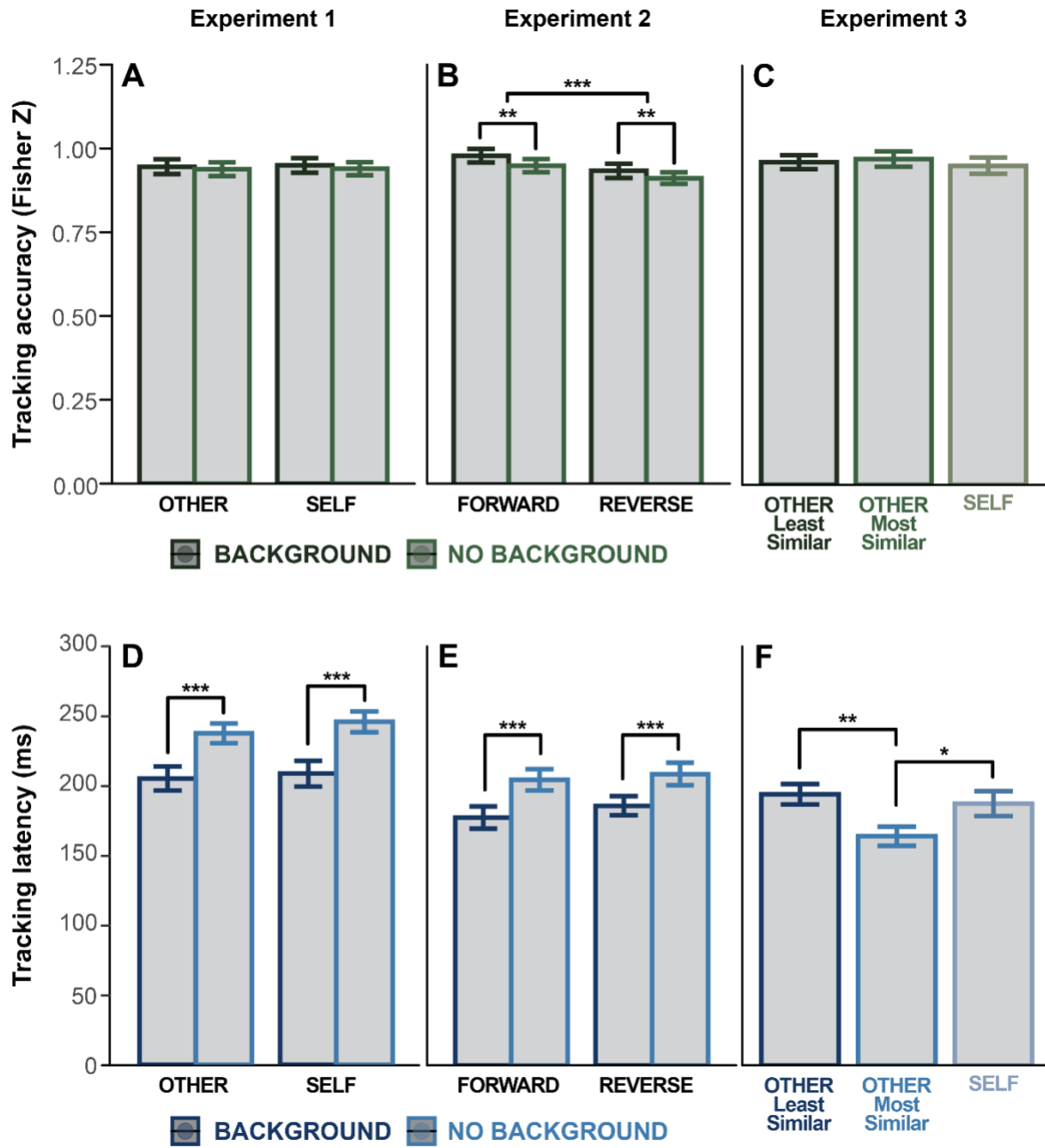
839   022-02506-2

840   Foulsham, T., & Kingstone, A. (2013). Where have eye been? Observers can recognise their own fixations.

841   *Perception*, *42*(10), 1085–1089. https://doi.org/10.1068/p7562

842   Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual

843   search. *Visual Cognition*, *14*(4–8), 704–715. https://doi.org/10.1080/13506280500193719

844   Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and

845   metacognitive processes while using computer-based learning environments. *Educational*

846   *Psychologist*, *45*(4), 203–209. https://doi.org/10.1080/00461520.2010.515935

847   Henderson, J. M., & Luke, S. G. (2014). Stable individual differences in saccadic eye movements during

848   reading, pseudoreading, scene viewing, and scene search. *Journal of Experimental Psychology:*

849   *Human Perception and Performance*, *40*(4), 1390–1400. https://doi.org/10.1037/a0036330

850   Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual

851   attention. *Vision Research*, *40*(10–12), 1489–1506. https://doi.org/10.1016/s0042-6989(99)00163-7

852   Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature*, *334*(6181), 430–431.

853   https://doi.org/10.1038/334430a0

854   Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in

855   psychtoolbox-3. *Perception*, *36*(14), 1–16.

856   Kok, E. M., Aizenman, A. M., Võ, M. L.-H., & Wolfe, J. M. (2017). Even if I showed you where you

857   looked, remembering where you just looked is hard. *Journal of Vision*, *17*(12), 2.

858   https://doi.org/10.1167/17.12.2

859   Marti, S., Bayet, L., & Dehaene, S. (2015). Subjective report of eye fixations during serial search.

860   *Consciousness and Cognition*, *33*, 1–15. https://doi.org/10.1016/j.concog.2014.11.007

861 Nelson, T. O. (1990). Metamemory: A Theoretical Framework and New Findings. In G. H. Bower (Ed.),

862   *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). Academic Press.

863   https://doi.org/10.1016/S0079-7421(08)60053-5

864 Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into

865   movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

866 Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two

867   approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.

868   https://doi.org/10.1016/j.tics.2006.03.006

869 Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control*

870   *of Language Processes*, *32*, 531–556.

871 Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information

872   processing tasks: Individual differences and cultural effects. *Vision Research*, *47*(21), 2714–2726.

873   https://doi.org/10.1016/j.visres.2007.05.007

874 Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). Use of eye movements

875   as feedforward training for a synthetic aircraft inspection task. *Proceedings of the SIGCHI*

876   *Conference on Human Factors in Computing Systems*, 141–149.

877   https://doi.org/10.1145/1054972.1054993

878 Sznabel, D., Land, R., Kopp, B., & Kral, A. (2023). The relation between implicit statistical learning and

879   proactivity as revealed by EEG. *Scientific Reports*, *13*(1), 15787. https://doi.org/10.1038/s41598-

880   023-42116-y

881 Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our Eyes Do Not Always Go Where We

882   Want Them to Go: Capture of the Eyes by New Objects. *Psychological Science*, *9*(5), 379–385.

883   Theeuwes, J., Kramer, A. F., Hahn, S., Irwin, D. E., & Zelinsky, G. J. (1999). Influence of attentional capture

884   on oculomotor control. *Journal of Experimental Psychology. Human Perception and Performance*,

885   *25*(6), 1595–1608. https://doi.org/10.1037//0096-1523.25.6.1595

886   van Gog, T., & Jarodzka, H. (2013). Eye Tracking as a Tool to Study and Enhance Cognitive and

887   Metacognitive Processes in Computer-Based Learning Environments. In *International Handbook*

888   *of Metacognition and Learning Technologies* (pp. 143–156). Springer, New York, NY.

889   https://doi.org/10.1007/978-1-4419-5546-3_10

890   Van Wermeskerken, M., Litchfield, D., & Van Gog, T. (2018). What Am I Looking at? Interpreting

891   Dynamic and Static Gaze Displays. *Cognitive Science*, *42*(1), 220–252.

892   https://doi.org/10.1111/cogs.12484

893   Veríssimo, I. S., Hölsken, S., & Olivers, C. N. L. (2021). Individual differences in crowding predict visual

894   search performance. *Journal of Vision*, *21*(5), 29. https://doi.org/10.1167/jov.21.5.29

895   Võ, M. L.-H., Aizenman, A. M., & Wolfe, J. M. (2016). You think you know where you looked? You better

896   look again. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10),
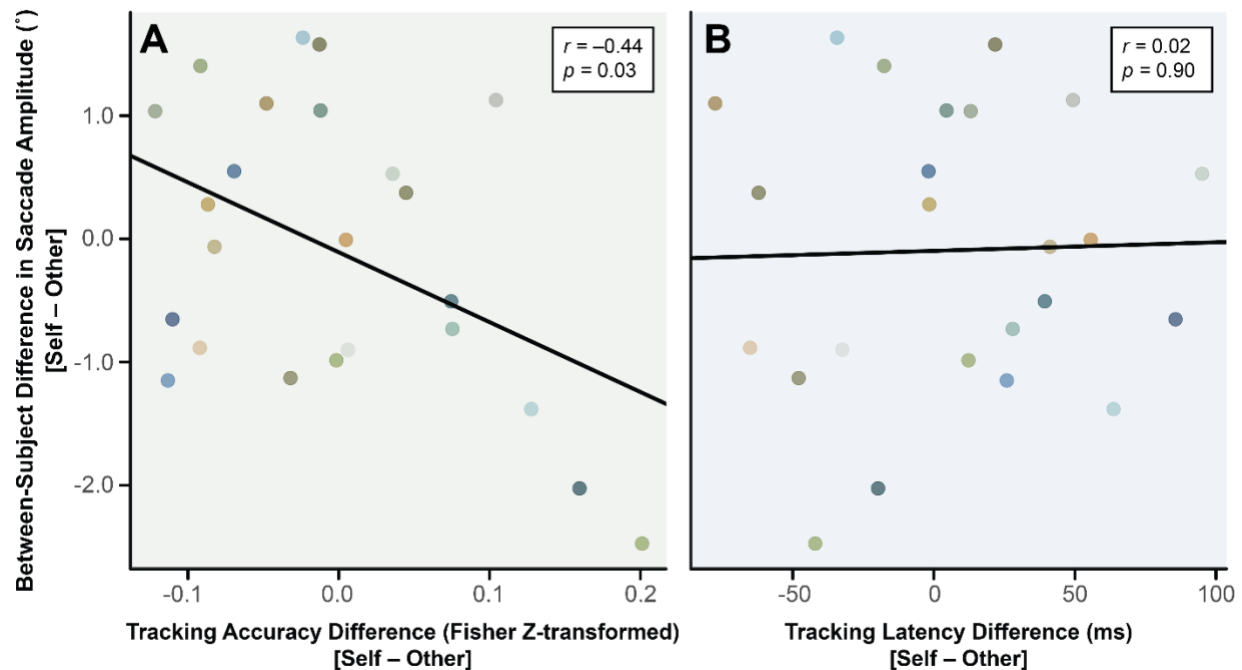
897   1477–1481. https://doi.org/10.1037/xhp0000264

Supplemental Materials

900   **Supplemental Figure 1. Tracking performance analyzed using the full length of trials equal**
901   **to or greater than 4000ms, across all three experiements.** The entire length of the trial was
902   included, as opposed to truncating each trial from 500 ms to 4000 ms, as done in the main analysis.
903   **(A-C)** Average tracking accuracy in each condition across all three experiments, calculated as the
904   peak of the cross-correlation (Fisher Z-transformed correlation value). Each bar graph represents
905   the results from a different experiment. Error bars represent ± 1 standard error of the mean (SEM).
906   Asterisks represent results of the 2×2 repeated-measures ANOVA for Experiments 1 and 2, and a
907   one-way repeated-measures ANOVA for Experiment 3, ****p** < .01, *****p** < .001. **(D-F)** Average
908   tracking latency in each condition across all three experiments, calculated as the lag corresponding
909   to the peak accuracy (i.e., the point on the x-axis of the CCG corresponding to the peak). Each bar
910   graph represents the results from a different experiment. Error bars represent ± 1 standard error of
911   the mean (SEM). Asterisks represent results of the 2×2 repeated-measures ANOVA for
912   Experiments 1 and 2, and a one-way repeated-measures ANOVA for Experiment 3, ***p** < .05, ****p**
913   < .01, *****p** < .001.

914

915   **Supplemental Figure 2. Individual variability in median saccade amplitude and tracking**
916   **performance.** (A) Pearson's correlation*, r* and *p* values at the top-left corner, comparing *Self –*
917   *Other* difference in tracking accuracy (x-axis) for all Experiment 1 participants with the between-
918   subject difference in median saccade amplitude (y-axis). Negative values on the y-axis means that
919   the 'Other' participant had larger saccade amplitude, while positive values means that the current
920   participant (Self) had larger saccade amplitude. On the other hand, negative values on the x-axis
921   means that the current participant performed better on the Other trials, while positive values means
922   that they performed better on the Self trials. (B) Correlation between the *Self – Other* difference in
923   saccade amplitude and the *Self – Other* difference in tracking latency, following the same
924   conventions as panel A. Positive values on the x-axis indicate larger latencies for the Self
925   condition, while negative values indicate larger latencies for the Other condition.

926