



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/219140/>

Version: Accepted Version

---

**Article:**

Ueno, T. and Allen, R.J. (2025) Running after two hares in visual working memory: Exploring retrospective attention to multiple items using simulation, behavioral outcomes, and eye tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 51 (3). pp. 405-420. ISSN: 0096-1523

<https://doi.org/10.1037/xhp0001270>

---

This is an author produced version of an article published in *Journal of Experimental Psychology: Human Perception and Performance*, made available under the terms of the Creative Commons Attribution License (CC BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15

## Title

Running after Two Hares in Visual Working Memory: Exploring  
Retrospective Attention to Multiple Items Using Simulation, Behavioral  
Outcomes, and Eye-tracking.

Taiji Ueno<sup>1,\*</sup> and Richard J. Allen<sup>2</sup>

<sup>1</sup> Faculty of Human Sciences, Takachiho University, Japan.

<sup>2</sup> School of Psychology, University of Leeds

## Correspondence

\* TU is now based at Division of Psychology and Communication, School of Arts and  
Sciences, Tokyo Woman's Christian University, Tokyo, Japan.

E-mail: [taijiueno@lab.twcu.ac.jp](mailto:taijiueno@lab.twcu.ac.jp)

Postal address: 2-6-1, Zempukuji, Suginami-ku, Tokyo Woman's Christian University, Japan.

16

**Acknowledgement**

17 This work was supported by JSPS KAKENHI Grant Numbers JP20H04494, JP22H01073,  
18 and JP23H01053.

19

20

**Abstract**

21 Multi-item retro-cueing effects refer to better working memory performance for multiple  
22 items when they are cued after their offset, compared to a neutral condition in which all  
23 items are cued. However, several studies have reported boundary conditions, and findings  
24 have also sometimes failed to replicate. We hypothesized that a strategy to focus on only  
25 one of the cued items could possibly yield these inconsistent patterns. In Study 1, a Monte-  
26 Carlo simulation showed that randomly selecting one of the cued items as the focus in each  
27 trial increased the chance of obtaining significant ‘multi-item retro-cueing effects’ on the  
28 mean accuracy over the trials, providing an incorrect conclusion if interpreted as evidence  
29 for attending all the cued items. These high rates to obtain such data fit with inconsistent  
30 patterns in the literature. To try and circumvent this situation, we conducted two new  
31 experiments (Studies 2A and 2B) where participants were explicitly instructed to fixate their  
32 gaze on all the cued positions, verified through eye-tracking (Study 2B). These produced  
33 robust multi-item retro-cueing effects regardless of previously identified boundary  
34 conditions. Notably, gazes were clearly fixated to multiple cued positions within each trial.  
35 Nevertheless, simulation revealed that our accuracy patterns could also in principle be  
36 produced by single item enhancement on each trial. The present study forms the first step to  
37 disentangle overt gaze-based allocation of attention from single-item focusing strategies,

38 while also highlighting the need for improved methodologies to probe genuine multiplicity  
39 in working memory.

40

41 *Keywords:* visual working memory, retrospective attention, retro-cue, monte-carlo

42 simulation, eye-tracking

43

**Public Significance Statement**

44 This study explores how attention can be directed to improve memory performance. We show  
45 via simulation that previous studies may only provide weak evidence for the ability to focus  
46 attention on more than one item in working memory when they are no longer present in the  
47 environment. We further show that people do look towards the locations of all cued items and  
48 can benefit their memory, when instructed to do so. However, questions remain about  
49 whether this really does show a genuine ability to enhance multiple items at the same time in  
50 working memory.

51

52

53

54

55

## Introduction

56 Visual working memory (VWM) retains objects and operates on them over time  
57 courses of a few seconds. VWM is closely related to attention. For example, studies have  
58 demonstrated that both *pre-cueing* one of the to-be-tested items before the onset of the target  
59 array and *retro-cueing* after the offset enhance accuracy for the cued items more than a  
60 neutral condition where all items are cued (Makovski & Jiang, 2007). As Souza and Oberauer  
61 (2016) reviewed, research on retro-cueing effects allows us to investigate the nature and  
62 mechanisms involved in focusing attention to internal information in working memory  
63 (Oberauer & Hein, 2012). For example, one can investigate how information in working  
64 memory is selected, and whether capacity limitations in working memory can be mitigated  
65 against via application of retrospective attention. By increasing the number of retro-cues, one  
66 can further examine whether such functions of the focus of attention can be expanded to  
67 multiple items in working memory. Furthermore, if the focus of attention can be  
68 retrospectively allocated to multiple representations in working memory, then one can next  
69 ask whether multiple representations are focused simultaneously or sequentially (see, Ort &  
70 Olivers, 2020, for consideration of this issue in the visual search domain). In the current  
71 work, our main question examined whether retrospectively attending to multiple items  
72 enhances accuracy. As outcomes in the literature are mixed, we first explored a possible  
73 reason for these inconsistencies, and then aimed to collect potentially stronger evidence for

74 the effect through task instructions and by tracking eye movement. As part of this approach,  
75 we inspected fixation data to clarify whether multiple items were attended simultaneously or  
76 sequentially.

77 Does *multi-item retro-cueing* reliably enhance accuracy for cued items? Whilst multi-  
78 item pre-cueing robustly enhances memory performance for cued items, there is still dispute  
79 regarding the effects of multi-item retro-cueing. An initial study did not find a significant  
80 multi-item retro-cueing effect (Makovski & Jiang, 2007). Subsequent studies investigated this  
81 more extensively and proposed boundary conditions for multi-item retro-cueing effects.  
82 These include (a) contextual matching between the number of cued items and the number of  
83 the probed items (Matsukura & Vecera, 2015); (b) dividing multiple cues in different  
84 hemifields (Delvenne & Holt, 2012); (c) sequentially presenting multiple retro-cues (Li &  
85 Saiki, 2014); and (d) using a feature-based cue (i.e., non-spatial cue) rather than a  
86 spatial/symbolic cue when the cues items are spatially distant from each other (Heuer &  
87 Schubö, 2016). Revealing the boundary conditions of a phenomenon is a typical and useful  
88 way to advance science because it means theories are refined. However, from a different  
89 viewpoint, boundary conditions also add complexity to theories. Thus, this is a balancing  
90 issue in that more complex theories are welcome only when these are necessary to explain the  
91 phenomenon.

92 However, these boundary conditions do not consistently explain the existing data. For  
93 example, there are studies that found the multi-item retro-cueing effect even without

94 contextual matching (e.g., Heuer & Schubö, 2016), without splitting cued items into different  
95 hemifields (e.g., Heuer & Schubö, 2016; Matsukura & Vecera, 2015); and regardless of the  
96 spatial distance between cued positions (e.g., Matsukura & Vecera, 2015). Studies have also  
97 failed to find the effects of sequentially presented multi-cues on the second cued item (e.g.,  
98 Van Moorselaar et al., 2015), although Souza et al (2015) replicated the effects of  
99 sequentially presented multiple cues on both the first and the second (subsequent) items.  
100 Thus, there is still disagreement on whether multi-item retro-cues enhance memory  
101 performance, particularly when using simultaneous presentation of cues (Souza & Oberauer,  
102 2016).

103         It is unclear why such inconsistencies are observed regarding multi-item retro-cueing  
104 effects. Without a systematic reason that affects the presence/absence of effects in a  
105 *probabilistic* manner, it is difficult to explain the existing patterns in the literature. We  
106 hypothesized that one of the candidates for such a reason might be a *single-item focusing*  
107 *strategy*. Specifically, if the cued items are probed more frequently than other items (i.e., high  
108 cue validity), then focusing on only one of the multiple cued items inevitably increases the  
109 probability to achieve higher *mean accuracy over trials* than randomly focusing on one of all  
110 studied items (i.e., the neutral cue condition where no items are preferentially cued). Since it  
111 is probabilistic whether the focused single item is probed, it is also probabilistic whether each  
112 participant exhibits higher mean accuracy over trials in the multiple-cue condition than the  
113 neutral cue condition (see, Ort & Olivers, 2020, for consideration of this issue in the visual

114 search domain). As such, it is not surprising that there are mixed outcomes regarding the  
115 multi-item retro-cueing effect, as it would be inappropriate to argue for attention towards  
116 multiple items if only one of the cued items is strategically focused on in each trial.

117         It is therefore important to estimate the actual chance rates to obtain a statistically  
118 significant difference between mean accuracies when a single-item focusing strategy is  
119 employed. If such rates are low, then this strategy-based account cannot explain the  
120 inconsistencies. A promising approach to estimate this chance rate under a hypothetical  
121 scenario (i.e., only single item is focused) is via statistical simulation, such as the Monte-  
122 Carlo method. This was the aim of Study 1. This approach involves repeated sampling of  
123 correct/incorrect responses from binomial distributions, whose parameters are derived from  
124 published studies. The rates of detecting a significantly higher mean accuracy in the multi-  
125 item retro-cueing condition can then be calculated in scenarios where only one item was  
126 focused on rather than attending multiple items. To foreshadow the outcomes, we found that  
127 such chance rates were high when the simulation parameters were taken from the existing  
128 literature that found multi-item retro cueing effects. In other words, evidence for  
129 ‘retrospectively attending multiple items’ may be weak even if the statistically significant  
130 multi-item retro-cueing effect was obtained. In addition, it is important to note that the  
131 estimated chance rates in Study 1 were not 100%. This means that when only a single-item  
132 was strategically focused, there is also a chance to fail to obtain a statistically significant

133 multi-item retro-cueing effect even if a genuine multi-item retro-cueing effect exists in  
134 human visual working memory.

135 To circumvent these situations, it is important to investigate multi-item retro-cueing  
136 under a situation where participants are actively discouraged from using a single-item  
137 focusing strategy. For this aim, we conducted two new experiments (Study 2A and Study 2B)  
138 in which participants were instructed to fixate their gazes to all the spatial positions of the  
139 cued items during the retention interval. Moreover, to collect stronger evidence, gaze was  
140 monitored in Study 2B by using an eye-tracker. Fixating gazes across all the cued positions  
141 would be a clear indicator that participants are not simply adopting a single-item focusing  
142 strategy. The rationale behind this approach is based on an active and facilitatory role of gaze  
143 position in memory after the offset of the studied items (Ferreira et al., 2008; Johansson &  
144 Johansson, 2014; Laeng et al., 2014; van Ede et al., 2019). Although Loaiza and Souza  
145 (2022) did not find a relationship between participants' spontaneous fixation duration during  
146 maintenance and recall accuracy in the fixated position, they did find a positive relationship  
147 when eye gaze to the *sequentially-presented* cued positions was explicitly instructed (see  
148 their Study 1B). Thus, this study justifies using instructed gaze fixation as an approach to  
149 attend to the multiple cued positions. We apply this technique to the simultaneously-  
150 presented multi-item retro-cueing paradigm. If retrospectively attending to multiple items is  
151 possible, then memory performance should be improved in the multi-item retro-cueing  
152 condition. Moreover, we may glean further insights on the mechanism of the multi-item

153 retro-cueing effect (i.e., sequential vs. simultaneous processing) by comparing fixation  
154 durations across the multiple cued positions.

155

## 156 **Study 1: Monte-Carlo Simulation**

### 157 **Transparency and Openness**

158 In Study 1, we used statistical R (R Core Team, 2023) for the simulation. All the  
159 simulation codes to reproduce the simulation data are available online  
160 (DOI 10.17605/OSF.IO/DRBXT). The detailed instruction for running the simulation is also  
161 provided online. The general overview of the simulation is provided in the next section.

162

### 163 **Aim and Overview of the Simulation**

164 A promising approach to estimate the chance rates of obtaining statistical significance  
165 under these hypothetical scenarios is via the Monte-Carlo method of statistical simulation.  
166 This was implemented in Study 1. In each simulation trial, we can calculate the probability  
167 for a focused item to be probed by chance under the assumption that *only one* of the cued-  
168 items was randomly selected as a to-be-focused item. This probability can be precisely

169 calculated based on number of items, number of cues, and cue validity. Then, the correctness  
170 (binary: correct or incorrect) in each trial can be simply simulated by sampling from a  
171 binomial distribution. The probability for sampling a correct case can be set to be higher  
172 when the randomly selected focused-item is probed by chance than when the focused-item is  
173 not probed by chance. The exact probabilities for these binomial distributions can be  
174 determined in an objective manner, such that the effect size of single-item retro-cueing (i.e.,  
175 single retro-cue condition vs. the neutral retro-cues condition) in the simulation matches with  
176 the real human data (later in detail). Once these probabilities for the binomial distributions  
177 are determined, then the correctness of each trial can be sampled regardless of the number of  
178 cues (2 or larger). Then, after repeating this sampling procedure for the number of trials in  
179 each condition and for the number of participants, the resultant data matrix can be submitted  
180 to a conventional statistical test (e.g., *t*-test or ANOVA) to examine a multi-item retro-cueing  
181 effect. Note that this is a simulation of a scenario where only one of the cued items is  
182 focused. Thus, even if significantly higher accuracy in the multi-item retro-cueing condition  
183 than the neutral-cues condition is detected, it does not provide evidence for successful  
184 retrospective attention to multiple items. By reiterating the same procedure many times (i.e.,  
185 Monte-Carlo method), we can estimate the probability of yielding statistical significance. The  
186 parameter values (i.e., number of items/cues/participants, cue validity, etc.) were taken from  
187 the existing literature that found a multi-item retro-cueing effect. We focused on the articles  
188 that reported both single-item and multi-item retro-cueing effects, as these are both necessary

189 for our simulation (see below). Moreover, we focused on the cases where multiple cues were  
190 presented simultaneously, because sequential presentation of multiple cues appears to have a  
191 reliable effect (see, Souza & Oberauer, 2016, for a review). In addition, we also simulated the  
192 other scenarios, where cued-validity is not 100% (e.g., Li & Saiki, 2014), and where a multi-  
193 item retro-cueing effect is modulated by another factor (e.g., Heuer & Schubö, 2016). The  
194 details will be explained later.

195

196

197 **Methods****Table 1.***Parameters from Each Human Experiments and Simulation Parameters*

Parameters and variables	Articles			
	Makovski & Jiang (2007)	Li & Saiki (2014)	Matsukura & Vecera (2015)	Heuer & Schubö (2016)
Number of participants	20~23	16~18	16~32	17~23
Number of studied items	6	4	6	4* <sup>1</sup>
Number of retro-cues	1, 2, 3, 4, 5* <sup>2</sup>	1, 2	1, 3	2* <sup>3</sup>
Number of neutrally-cued items	6	4	6	4
Number of probes	1	1	1, 3* <sup>4</sup>	1
Number of trials	48~90	48* <sup>5</sup>	70	192
Single-cueing effect size				
Hedge's <i>g</i>	0.611	1.433* <sup>6</sup>	2.067	1.105* <sup>7</sup>
Simulation ( $p_{unfocus}$ )* <sup>8</sup>	0.600	0.600	0.600	0.600
Simulation ( $p_{focus}$ )* <sup>9</sup>	0.639	0.726	0.738	0.649

*Note.* The *N* of participants/trials varied depending on the experiments within each article. Thus, we adopted the largest value for our simulation; \*1 There were 8 items in the study array, but only half of them (either hemifield) was task-relevant; \*2 We adopted 2-item retro-cueing condition for our simulation; \*3 There was not a single-cueing condition; \*4 They employed a set-probe testing procedure; \*5 There were 144 trials in the double-cue condition, but there were 48 trials where each one of the cued positions were probed; \*6 The only statistics for the single-cueing effect reported in Li and Saiki (2014) was the contrast between the 'withdrawal condition' and the neural cue condition. Thus, we used these statistics in order to estimate the single-cueing effect size; \*7 This value was conservatively taken from the double-cueing effect (see main text); \*8 The simulation parameter was set to be 0.6. This was set arbitrarily but this absolute value is not important. What matters is the relative difference from the  $p_{focus}$  value; \*9 These values were determined by a grid search so that the single-cueing effect size matched with human data (see main text).

198 *Parameters from the Target Articles*

199           The target articles that we simulated are listed in Table 1. The parameters ( $N$  of  
 200 participants, trials, cues, targets, and probes) were taken from each article. When multiple  
 201 experiments were conducted in each article, the largest value was used. The effect size index  
 202 (Hedges'  $g$ ) for the single-item retro-cueing effect was converted from the paired  $t$ -statistics  
 203 (single-cue condition vs. neutral-cue condition) reported in each paper.

204

### 205 ***Simulation Dataset and Parameters from the Human Experiments***

206           First, we specified the model to be studied using the Monte Carlo simulation. In each  
 207 trial, only one of the cued items were randomly selected as a to-be-focused item. Thus, the  
 208 probability to select the  $i$ -th item ( $p_{select}$ ) simply follows a discrete uniform distribution:

209

$$210 \quad p_{select}(X = i\_th\ item) = \frac{1}{N_c} \quad (i = 1, 2, \dots, N_c) \dots Equation\ (1)$$

211

212 , where  $N_c$  is the number of cues. In the neutral cue condition (i.e., control condition),  $N_c$  was  
 213 equal to the number of the studied items.

214           Next, we considered the probability for each item to be probed in the test phase. In a  
 215 single-probe change-detection paradigm, the probability for each item to be probed depends

216 on the number of cues ( $N_c$ ) and the cue-validity level (i.e., the degree of predictiveness of the  
 217 cues). We first simulated the case of the maximum cue-validity level (i.e., 100%), and  
 218 consider lower validity cases later. When cue-validity is 100%, then the probability for one  
 219 ( $j$ -th item) of the cued items to be probed ( $p_{probe}$ ) follows a discrete uniform distribution as  
 220 follows:

221

$$222 \quad p_{probe}(X = j\_th\ item) = \frac{1}{N_c} \quad (j = 1, 2, \dots, N_c) \dots Equation (2)$$

223

224 Next, when the probed item (or the lure probe in the multiple-probes situation) was  
 225 coincidentally from the same position as the focused (single) item, then it is natural to expect  
 226 a higher accuracy than when these were different. Then, the correctness in the  $k$ -th trial ( $C_k$ )  
 227 can be simply simulated by sampling from the binominal distribution as follows:

$$228 \quad C_k \sim Binomial(n$$

$$229 \quad = 1, p) \begin{cases} p = p_{focus}, \text{ when a probed item is taken from the focused position} \\ p = p_{unfocus}, \text{ when a probed item is not taken from the focused position} \end{cases}$$

230

... Equation (3)

231 , where the probability for the binomial distribution ( $p_{focus}$ ) was set to be higher than  
232 ( $p_{unfocus}$ ). The simplicity of this equation as a model of human recognition is discussed later  
233 in the results section. The specific values for the probability parameters were determined  
234 objectively as follows: First, it is entirely reasonable to assume that participants take this  
235 single-item focusing strategy when the number of cues is one (i.e., single-item retro-cueing  
236 condition). Then, the specific *difference* between ( $p_{focus}$ ) and ( $p_{unfocus}$ ) values can be  
237 determined so that the effect size of the single-item retro-cueing effect in the Monte Carlo  
238 simulation matches that in the real human data. Taking Makovski and Jiang (2007) as an  
239 example, the single-item retro-cueing effect sizes (i.e., single cue condition vs. neutral cues  
240 condition) in their experiments were Hedges'  $g$  of 0.655, 0.470, 0.530, and 0.930  
241 (Experiments 1a, 1b, 2a, 2b, respectively). A random-effect meta-analysis (ESCI software:  
242 Cumming, 2012) can aggregate these effect sizes to estimate a single score, Hedges'  $g$  =  
243 0.611. Then, bearing this integrated single-cueing effect size in mind, we conducted a grid  
244 search to determine the ( $p_{focus}$ ) and ( $p_{unfocus}$ ) parameter values for simulating Makovski  
245 and Jiang (2007)'s study. More specifically, we first fixed the ( $p_{unfocus}$ ) parameter to 0.600  
246 whilst gradually increasing the ( $p_{focus}$ ) parameter by 0.001. Then, the correctness in the  $k$ -th  
247 trial ( $C_k$ ) of the single-cue condition in the simulation was sampled by the Equation (3) above  
248 for the same number of trials/participants as Makovski and Jiang (2007). This allows to  
249 compute the single-cueing effect size in the simulation under a given ( $p_{focus}$ ) parameter  
250 value. Once the resultant effect size in the simulation reached Hedges'  $g$  of 0.611 (i.e., real

251 human data), then we stopped increasing the ( $p_{focus}$ ) parameter value. In case of Monte-  
252 Carlo simulation for Makovski and Jiang (2007) study, the ( $p_{focus}$ ) was determined to 0.639.  
253 The same procedure was taken in determining the ( $p_{focus}$ ) parameter value in simulating  
254 each study listed in Table 1.

255         Next, the obtained ( $p_{focus}$ ) parameter values can be used in sampling the correctness  
256 value in the multi-cue trials and in the neutral-cue trials of the simulation as well (Equation  
257 3). This is a plausible procedure because the current Monte-Carlo approach assumes the  
258 scenario where a single-item focusing strategy is taken in the multi-cue trials. By reiterating  
259 the sampling of the correctness value in the  $k$ -th trial ( $C_k$ ) for the same number of  
260 trials/participants as the existing literature (see Table 1), we can generate the simulation data  
261 matrix under an assumed hypothetical scenario. Then, a paired  $t$ -test can be conducted to  
262 investigate whether a statistically significant difference is detected between the multi-item  
263 cueing condition ( $1 < N$  of cues  $< N$  of studies items) and the neutral condition ( $N$  of cues =  $N$   
264 of studied items). If statistical significance is observed, it would be inappropriate to  
265 confidently interpret this as evidence for attending all cued items. Instead, such a result  
266 should be interpreted as supporting evidence for the argument that one can obtain  
267 significantly higher mean accuracy in the multi-cued condition than in the neutral-cues  
268 condition even when just one of the cued items is strategically focused on. Finally, all the  
269 procedures so far can be repeated multiple times in Monte-Carlo simulation (100 times in our  
270 case) so that the chance rates to observe statistical significance can be calculated in each

271 experimental situation of the existing literature. The number of the iterations was determined  
272 by conducting a self-replication. A higher number of iterations leads to more stable  
273 estimations (i.e., less affected by sampling variations) but requires more time to conduct. The  
274 effect size of sampling variances on the outcome of the model can be evaluated empirically  
275 by re-running the simulation. We re-conducted the simulation with the number of iterations  
276 as 500, and the outcomes were very similar (e.g., the difference in the estimated chance rates  
277 was less than 5%). Thus, for the sake of efficiency we report analyses using 100 iterations.

278         We note here that Heuer and Schubö (2016) did not include the single-item retro-  
279 cueing condition, in which case we computed the effect size of the multi-item (num of cues =  
280 2 in this case) retro-cueing effect and used it as a conservative index for the single-item  
281 cueing effect size of the human participants. Of course, the real single-item retro-cueing  
282 effect size should be larger than the multi-item retro-cueing effect size. Thus, we obviously  
283 underestimated the size of the single-item retro-cueing effect size in this case, which is why  
284 our approach is conservative. The larger the single-item retro-cueing effect size, the higher  
285 the chance rates to detect the multi-item retro-cueing effect in the current Monte-Carlo  
286 simulation. In other words, underestimating the single-cueing effect size inevitably  
287 underestimates the chance rates to obtain statistical significance in simulation. Then, if the  
288 resultant chance rates are still found to be higher in such a conservative situation, then we can  
289 confidently argue that the rates of obtaining statistical significance in the experimental  
290 situation are much higher than our conservative estimate.

291           Finally, although Delvenne and Holt (2012) was one of the articles that argued to find  
292 the boundary condition for multi-item retro-cueing effect, we could not simulate their  
293 experimental situation because the detailed statistics for neither the single-item retro-cueing  
294 effect nor the multi-item retro-cueing effect were available (i.e., only  $p$  value was reported as  
295  $p < .05$ ).

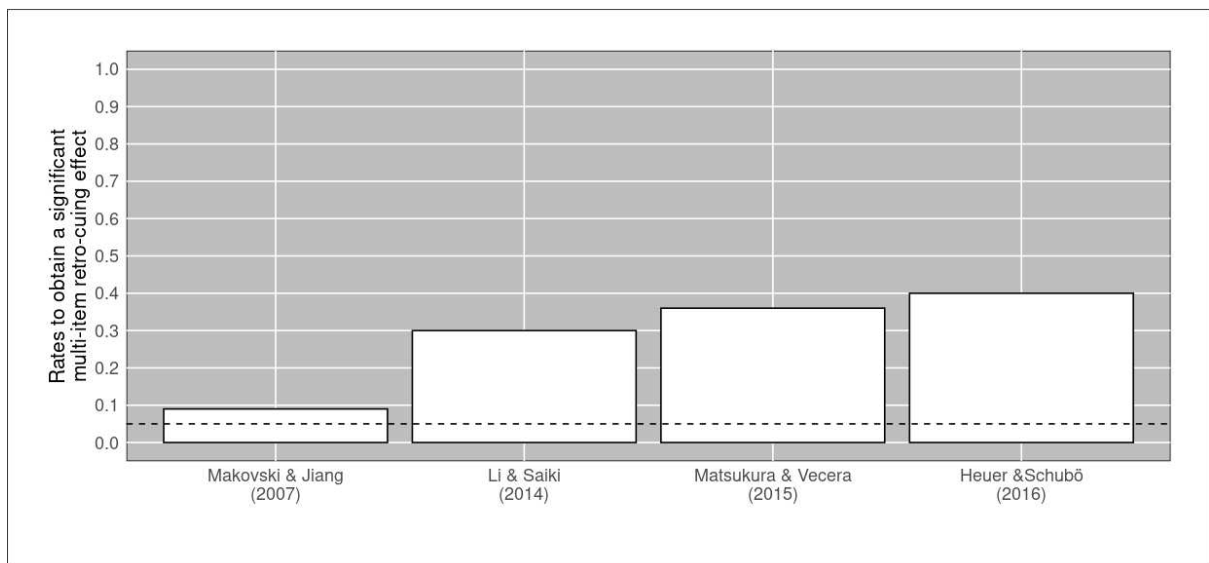
296

## 297 **Results and Discussion**

298

299 **Figure 1**

300 *Rates for Obtaining a Statistically Significant Difference in Accuracy between the Multi-item*  
 301 *Retro-cueing Condition and the Neutral-cueing Condition in the Monte-Carlo Simulation.*



302

303 *Note.* A dashed horizontal line indicates  $y = 0.05$  (i.e., conventional alpha level)

304

305 Figure 1 shows the outcomes of the Monte-Carlo simulation. Specifically, the Y-axis  
 306 is the number of the Monte-Carlo iterations that detected statistical significance between the  
 307 multi-item retro-cueing condition and the neutral-cues condition where all the items are cued,  
 308 divided by the total number of the simulation run (i.e., 100) in each scenario. Note that only  
 309 one of the cued items was strategically focused on in this simulation. Thus, the y-axis shows

310 the chance rates to draw an incorrect interpretation that participants were successfully  
311 attending to multiple items.

312 First, when simulating the initial study by Makovski and Jiang (2007), who did not  
313 find a significant multi-cueing effect, the estimated chance rates for misinterpretation were  
314 close to 5%. In contrast, when simulating the other studies, which found significant multi-  
315 cueing effects, the obtained chance rates for misinterpretation were much higher (30% ~  
316 40%) than the conventional alpha level (5%). This means that in the latter cases, one can find  
317 a statistically significant difference between the multi-cue and neutral-cue conditions at a  
318 chance rate of higher than 5% even if participants focused on only one of the cued items. Of  
319 course, we cannot confidently assert whether each participant in these studies covertly  
320 applied such a strategy. However, such higher chance rates are consistent with the fact that  
321 some studies obtained a statistically significant effect whilst others did not. The crucial  
322 difference between Makovski and Jiang (2007) and the other studies was the effect size of the  
323 single-item retro-cueing effect (i.e., single cue vs. neutral cues). As Table 1 shows, this effect  
324 size was much smaller in Makovski and Jiang (2007), meaning it was unlikely to generate a  
325 statistically significant multi-item retro-cueing effect even if a single-item focusing strategy  
326 was adopted in the multi-cues condition. The larger effect size in the remainder of the studies  
327 means there is more scope for a single-item focusing strategy to explain the higher accuracy  
328 observed in the multi-cue conditions.

329           Before simulating other scenarios, it is worthy to discuss whether the simplification of  
330 our model is justified, particularly regarding Equation (3). Specifically, we sampled the  
331 correctness value in the  $k$ -th trial ( $C_k$ ) from the simple binomial distributions. One may  
332 justifiably criticize this approach in that various processes during the yes/no recognition  
333 judgment are assumed to be a black box in the current simulation. We acknowledge that our  
334 model did not specify the details of these processes in Equation (3) but argue that the output  
335 of the black box is correct. This is because we determined the probabilities of the binomial  
336 distribution (i.e.,  $p_{focus}$ , and  $p_{unfocus}$ ) based on the single-cueing effect size of the real  
337 human data in each study. In a single-cue trial, it is reasonable for a participant to use a  
338 single-item focusing strategy. Thus, as far as the outputs of the black box correctly simulate  
339 the single-cueing effect size of real participants, we can safely argue that the outcomes of the  
340 black box are close to that of the human outputs who employ a single-item focusing strategy.  
341 In a similar vein, we can justify our use of Equations 1-3 in simulating the set-probe testing  
342 method (i.e., number of probes  $> 1$ , see, Matsukura & Vecera, 2015). We acknowledge that  
343 the detailed processes should be different between a single-probe test and a set-probe test.  
344 However, once again, it is reasonable to assume that participants would naturally use a  
345 single-item focusing strategy in a single-cue trial even if a set-probe testing method is  
346 employed (Matsukura & Vecera, 2015). Then, as far as the probabilities of the binomial  
347 distributions (i.e.,  $p_{focus}$ , and  $p_{unfocus}$ ) are determined on the basis of the single-cueing  
348 effect of the participants from Matsukura and Vecera (2015), we can safely argue that the

349 outcomes of Equation (3) reflect the real underlying (hidden) processes of single-item  
350 focusing in a set-probe test.

351

### 352 *Lower Cue-Validity Level*

353 Can this strategy-based account hold even if the cue-validity level is low? The  
354 simulation above assumed the case of the maximum cue-validity (100%). Therefore, the  
355 probability for one ( $j$ -th item) of the cued items to be probed ( $p_{probe}$ ) was equal to one  
356 divided by the number of cues (Equation 2). However, some studies have investigated cueing  
357 effects when the cued items did not have a higher chance to be probed than the un-cued items  
358 (e.g., Experiment 3 of Li & Saiki, 2014). When simulating such a situation, one may think  
359 that Equation (2) should be changed to one divided by the number of studied items, rather  
360 than the number of cues. As a result, the simulation outcome would also change. However,  
361 we argue that the simulation outcome does not change. The reason lies in the accuracy  
362 scoring method in the multi-cued condition from the human experiments. Specifically, mean  
363 accuracy in the multi-cue condition per participant is calculated only from the cue-valid trials,  
364 where the cued item(s) are probed. In contrast, when an un-cued item is probed, the response  
365 in such a cue-invalid trial is not included when computing the mean accuracy of a participant  
366 in the multi-cued condition. This is a standard scoring method especially when contrasting

367 accuracy in the multi-cue and neutral-cue conditions (e.g., Li & Saiki, 2014). As far as this  
368 standard scoring method is taken, Equation (2) holds without an amendment. This is because  
369 in the cue-valid trials, the chance for a single focused-item to be probed is equal to one  
370 divided by the number of cues. As we already showed in the simulation of Li and Saiki  
371 (2014) in Figure 1, the chance rate to obtain a statistically significant multi-cueing effect was  
372 much higher than 5%. In other words, a single-item focusing strategy is still an effective  
373 strategy to increase the accuracy in the multi-cues condition than the neutral-cues condition  
374 even when cue validity is low.

375

### 376 *Interaction as a Counterargument for a Single-Item Focusing Strategy?*

377 Finally, one can argue that boundary conditions (i.e., interactions with another factor)  
378 are incompatible with the single-item focusing strategy account. Specifically, if participants  
379 focused on only one of the cued items, then the conditions of another factor such as spatial  
380 positions/distances between cues should not have made any difference for the multi-item  
381 cueing effect (see, Heuer & Schubö, 2016, for such an explanation). However, if this strategy  
382 results in a statistically significant difference being obtained in a probabilistic manner, then it  
383 is also probabilistic to obtain a significant difference in one condition whilst not to obtain it in  
384 the other condition (i.e., the interaction). Thus, it is necessary to demonstrate actual

385 probabilities to obtain such a discrepancy under the scenario when a single-item focusing  
386 strategy is employed.

387         The Monte-Carlo simulation strategy is as follows: We reiterated the abovementioned  
388 simulation of Heuer & Schubö (2016) once again, but this time the multiple-cue trials were  
389 divided (i.e., into two ‘conditions’) simply in terms of whether each trial was odd-numbered  
390 or even-numbered. Then, accuracy in each condition was compared to that in the neutral-cue  
391 condition. If a significant difference from the neutral-cue condition is detected in one  
392 condition whilst not in the other condition, then such a dissociation is regarded as an  
393 interaction. After repeating the same procedure 100 times, we calculated the rates for such a  
394 dissociation.

395         As a result, the rates for obtaining a dissociation was 0.3 (30%). We simulated the  
396 situation where a single-item focusing strategy was adopted, and the double-cue trials were  
397 randomly divided into two. As a result, there was a higher chance than 5% to obtain an  
398 interaction. Since both a focused item and a probed item were selected randomly from the  
399 cued items, it was also probabilistic either to obtain a significantly higher accuracy than the  
400 neutral-cues condition or not to obtain it.

401         Taken together, when interpreting a statistically significant multi-item retro-cueing  
402 effect, it is important to consider the extent to which a single-item focusing strategy can  
403 account for the outcome. The crucial factor is the effect size of the single-item retro-cueing

404 effect. If it is too high, one needs to be cautious. Of course, there is no evidence to conclude  
405 that participants in past studies covertly applied such a strategy. It is also important to note  
406 that we are not arguing that genuine multi-item cueing is not possible. Instead, stronger  
407 support for a significant multi-item retro-cueing effect may be obtained by increasing the  
408 likelihood for attentional allocation towards multiple cued items, and capturing evidence of  
409 participants' efforts to do so.

410         Therefore, to collect such evidence, we conducted two new experiments with human  
411 participants by attempting to control gaze positions during the maintenance phase.  
412 Specifically, during the whole maintenance phase, participants were instructed to fixate their  
413 gaze across all the spatial positions where the cues had appeared. We know that participants  
414 can follow instructions to strategically adjust their attentional focus between items in working  
415 memory tasks (e.g. Allen & Ueno, 2018; Atkinson et al., 2022; Allen et al., 2024 for a  
416 review). The rationale behind applying such an approach in the present study is based on an  
417 active and facilitatory role of gaze position in memory after the offset of the studied items  
418 (Ferreira et al., 2008; Johansson & Johansson, 2014; Laeng et al., 2014; van Ede et al., 2019).  
419 Relatedly, Loaiza and Souza (2022) found that instructed gaze fixation to the cued positions  
420 led to higher recall accuracy of the cued (i.e., gazed) position. Evidence of direction of gaze  
421 fixation across cued positions would be a clear indicator that participants followed the  
422 instruction to attend all multi-cued items. We therefore further conducted Study 2B to  
423 replicate outcomes from Study 2A, and also to collect direct evidence for gaze positions

424 through eyetracking. A significant multi-item retro-cueing effect in this context, along with  
425 evidence of gaze direction to multiple cued locations, would offer a stronger indication of  
426 real multi-item cueing effects than have been observed to this point.

427         Three points should be made regarding our gaze-based approach. First, this cannot  
428 firmly rule out the possibility that participants covertly focus attention on a single item,  
429 independent of gaze fixation. However, with evidence for gaze position, we can safely reject  
430 an account in which participants exclusively focused on only one of the multi-cued items.  
431 Namely, we can argue that at least an overt form (i.e., gaze-based) of attention is allocated to  
432 multiple representations. Second, gaze fixation at multiple positions does not necessarily  
433 mean that the representations from those positions are activated simultaneously in each trial  
434 (see, Orts & Olivers, 2020, for raising this issue in the visual search domain). Rather, the  
435 representation for each item would be re-activated (or refreshed) sequentially as gaze is  
436 fixated to each position. This sequentiality vs. simultaneity question will be discussed further  
437 after analysis of the eyetracking data. Third, and relatedly, while gaze fixation at multiple  
438 positions clearly rejects a deliberate single-item focusing strategy, it does not necessarily  
439 mean that both cued items are activated and enhanced by gaze-based attention during the  
440 retention interval. Instead, only one of the fixated items might actually benefit from gaze-  
441 based attention. This issue will be further discussed by examining the relationship between  
442 accuracy and fixation duration.

443

444

## Studies 2A and 2B: Human Experiments

### 445 Aim and Rationale

446 In Study 2A, we aimed to investigate the multi-item retro-cueing effect whilst  
447 explicitly instructing the participants to fixate their gaze across all the multiple cued-  
448 positions. In Study 2B, we aimed to replicate the findings of Study 2A as well as glean  
449 further evidence for fixating gazes across all the multiple spatial positions using eye-tracking.  
450 Our principal focus was on whether performance in the double-cue condition would be  
451 superior to that observed in neutral cue trials when a single-item focusing strategy, at least in  
452 the overt form (i.e., gazed-based attention), was explicitly discouraged. For the eye-tracking  
453 outcomes in Study 2B, we also examined whether gaze was directed to cued locations for a  
454 longer period than to uncued locations during double-cue trials.

455

### 456 Methods

457 *Transparency, Openness (Data Availability, Pre-registration), and Participants*

458           We report the sampling plan, all manipulations, and all measures in the study. All the  
459 data, analysis code, and research materials are available online  
460 (DOI 10.17605/OSF.IO/DRBXT). We used *R* (R Core Team, 2017) with *power.t.test*  
461 function to perform the power analysis. We preregistered the sampling plan and the analytic  
462 strategies of two studies beforehand (AsPredicted: [https://aspredicted.org/OIA\\_IZA](https://aspredicted.org/OIA_IZA)).  
463 However, for transparency, we note that the studies reported here were conducted in 2018,  
464 with subsequent delays arising due to COVID19 complications. In other words, the effect  
465 size used in the power analysis (i.e., within-group Cohen's *d* for the multi-cues vs. neutral-  
466 cues conditions) was determined based on the available literature at the time of pre-  
467 registration. Given the literature has developed since then (e.g., DiPuma et al., 2023), it is  
468 informative to conduct a sensitivity power analysis (e.g., Perugini et al., 2018) with the  
469 sample size that we collected. As a result, the minimum effect size that our study ( $N = 32$ )  
470 was sensitive to detect was Cohen's *d* of 0.51 (power = 0.8, alpha = 0.05). This effect size is  
471 common in the working memory literature. Thus, our original pre-registered sampling plan  
472 was appropriately powered to detect effect sizes of interest in these studies.

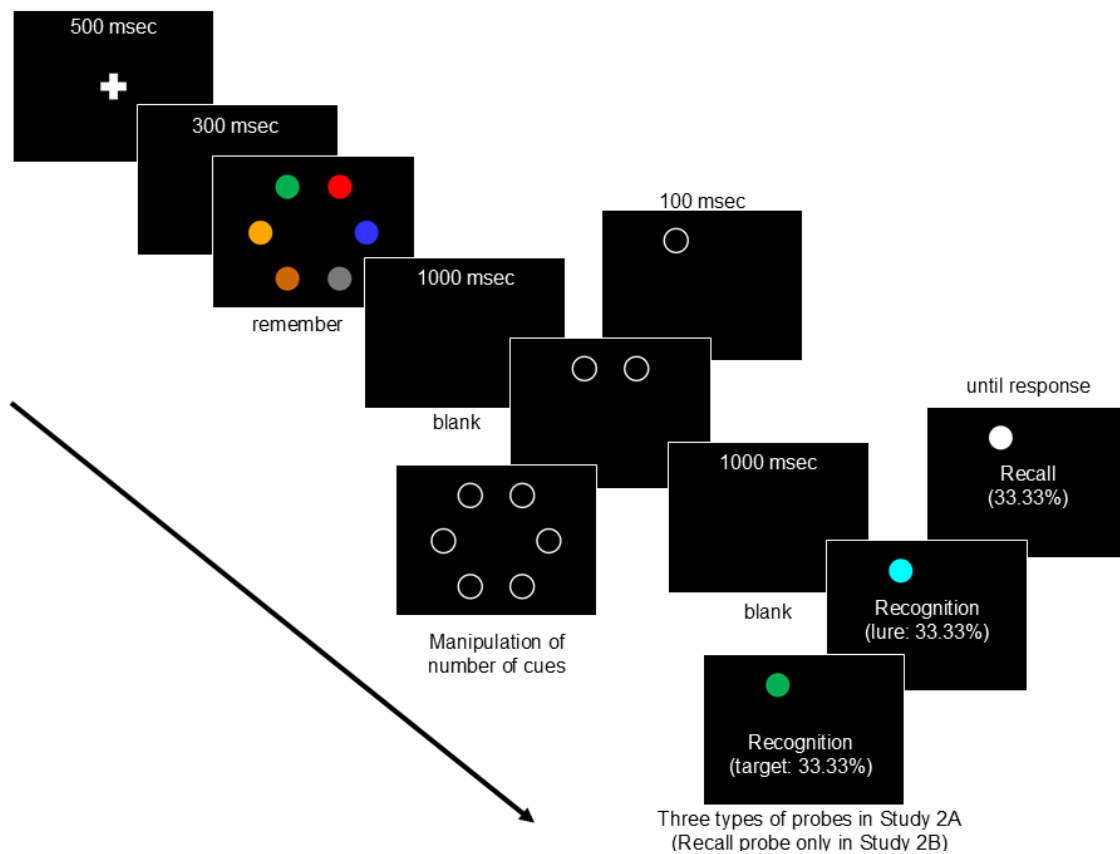
473           In each study, we continued data collection until 32 participants completed the  
474 experiment. There were 9 females and 23 males in Study 2A; 12 females and 20 males in  
475 Study 2B. The mean ages (and *SD*) were 19.15 (1.05) in Study 2A and 19.81 (1.06) in Study  
476 2B. None of these participants were excluded due to the pre-registered criteria: i.e.,  
477 participants whose mean response time (averaged over all the trials) was below 400ms; who

478 did not complete the study; who encountered a PC problem; and whose overall mean  
479 accuracy was 2.5 SD above/below the mean in each condition. We did not pre-register any  
480 trial-level exclusion criteria. Thus, all the data were submitted to the analysis. All the  
481 participants were from Takachiho University, Japan. They took part in the 45-minute  
482 experiment and were paid (1,000 Japanese Yen) for their participation. All had normal vision  
483 and discrimination ability for the colors. The experimental protocol had been submitted to the  
484 dean of Takachiho University in advance, who was also the chair of the ethical committee  
485 and approved of the protocols. All the participants gave written informed consent.

486

487 ***Design, Materials, and Procedure***

488

489 **Figure 2**490 *Flow of a Trial in Study 2A and Study 2B*

491

492 *Note. In the recognition trials (i.e., a colored circle probe), the participants made an*

493 *“old/new” judgment on the color of the target in that position via a keypress. In the recall*

494 *trials (a white circle probe), the participants selected the color of the target in that position*

495 *via a 9-alternative keypress (with different colored patches attached to each of 9 keys on the*

496 *keyboard).*

497

498           **Study 2A.** This experiment followed a repeated-measures design, with the number of  
499 retro-cues (one, two, or six) as the within-participant factor. Testing was controlled using a  
500 HSP3 (Hot Soup Processor, ver.3) program (<http://hsp.tv/>). The materials and the flow of a  
501 trials (Figure 2) were the same as Makovski and Jiang (2007), except for the longer duration  
502 of the maintenance phase after the offset of the retro-cues, and the testing method (see later).  
503 Each trial began with a warning cross (500ms), followed by a blank screen (300ms). Then, on  
504 each memory display, six coloured circles (diameter =  $1.31^\circ$ ) appeared for 1,000ms  
505 equidistantly on an invisible circle (diameter =  $9.84^\circ$ ), centred around the screen center. The  
506 background color was black. The colors of the six circles were randomly selected without  
507 replacement from *green* (RGB= 0, 255, 0), *red* (255, 0, 0), *blue* (0, 0, 255), *light blue* (0, 255,  
508 255), *purple* (255, 0, 255), *yellow* (255, 255, 0), *gray* (127.5, 127.5, 127.5), *brown* (204, 102,  
509 0), and *orange* (255, 165, 0). Before the experiment, all the participants were shown the nine  
510 coloured circles and confirmed that they were able to discriminate each color easily. After the  
511 offset of the studied items, a blank screen (1000ms) was inserted. Then, one, two, or six  
512 peripheral attentional cues appeared for 100ms in the form of open circles (diameter =  $1.31^\circ$ ,  
513 line-color = white, filled color = black). Each represented the single-cue condition, multiple-  
514 cues condition, and neutral-cues condition, respectively. The number of trials for each of  
515 three cueing conditions was 45 in Study 2A (135 trials in total). The spatial positions of the  
516 cues were randomly selected from the six studied items' positions. The instruction was  
517 crucial in this study: Participants were informed that only the cued locations would be

518 probed, and therefore they were instructed to focus on the items that had appeared at the  
519 retro-cued locations and to ignore the items at the un-cued locations. Thus, overall probability  
520 that a cued item would be probed was 100% (though this was obviously reduced for each  
521 individual item in the multi-cue conditions). Participants were also instructed that when two  
522 cues appeared, one of these cued positions was always probed, and therefore they should  
523 focus on *both* (this word was highlighted in red in the instruction screen) items that had  
524 appeared at the retro-cued locations. When all the six positions were cued (i.e., neutral-cues),  
525 the participants were instructed to focus on all the studied items as the cues did not predict  
526 the position of the probe. Moreover, the participants were instructed to fixate their gaze on all  
527 the cued positions during the maintenance phase (i.e., after the offset of cues and before the  
528 onset of a probe). To facilitate this gaze distribution, we made this duration longer (1,000ms)  
529 than Makovski and Jiang (2007), who set this duration at 400ms.

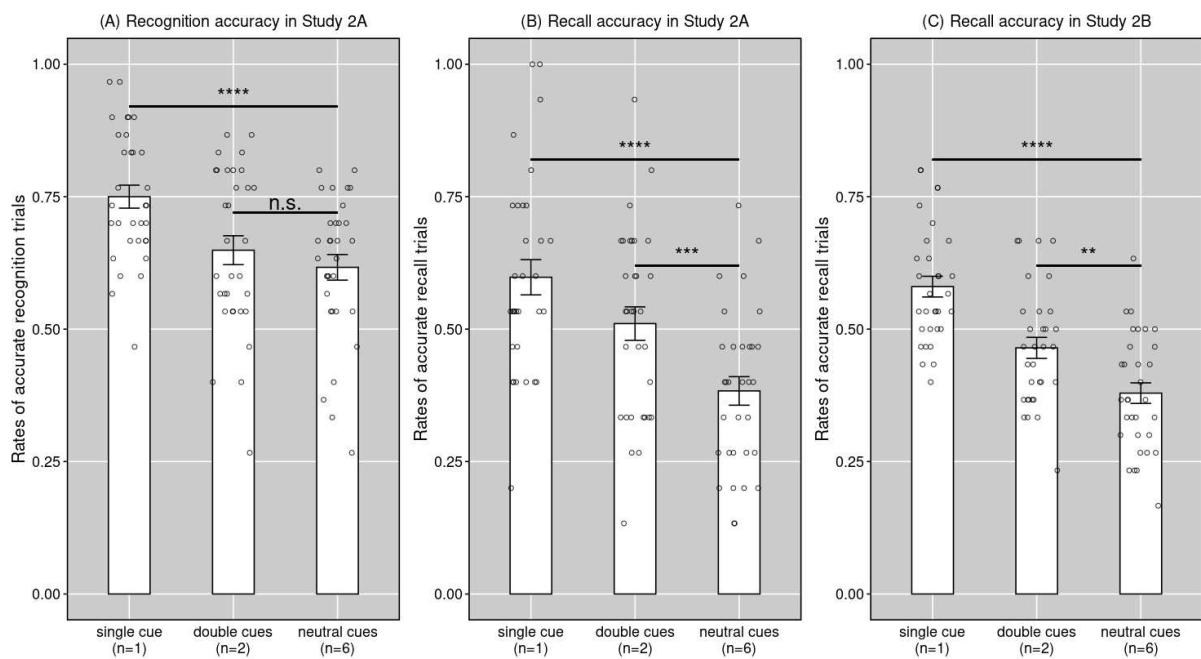
530         At the end of each trial, a single probe appeared at one of the cued position(s). Whilst  
531 Makovski and Jiang (2007) used a single-probe recognition paradigm, we added a recall  
532 probe as well, where a participant was required to recall the color in Study 2A. The response  
533 method involved selecting the color of the target item by a keypress. To aid this response,  
534 different colored patches were attached to each of 9 keys on the keyboard (“1” ~ “9”). This  
535 was done to address the issue of sensitivity. Compared to the 2-alternative choice in probe  
536 recognition, chance rate is much lower for recall from nine-possible colors. Therefore, there  
537 were three types of probes in the test phase: recall probe (i.e., a white, filled circle), a positive

538 recognition probe (i.e., a filled circle whose color was the same as the studied item in the  
539 probed position), and a negative recognition probe (i.e., a filled circle whose color was  
540 randomly selected from one of the three, un-presented colors at that trial). These three types  
541 of probes appeared with equal probability (i.e., 33.33% for recall and 66.66% for  
542 recognition), randomly intermixed within blocks. If a colored circle appeared as a probe, then  
543 participants were required to press “s” (same) or “d” (different). If a white, filled circle  
544 appeared as a probe, then the participants were required to select the color of the target item  
545 in that position by a keypress. The probe remained on screen until a keypress response was  
546 made. Speed was not required, and accuracy was emphasized. At every trial, participants  
547 were engaged in an articulatory suppression (saying ‘da’, ‘da’, ...) until a key press.

548         **Study 2B.** In Study 2B, we focused on the recall testing method to optimise  
549 sensitivity and avoid the higher chance rate of the recognition probe. Thus, all the probes  
550 were white, filled circles. The number of trials was 90 in total (i.e., 30 trials for each of the  
551 single-cue, multi-cues, and neutral-cues conditions). Moreover, in Study 2B, we measured the  
552 gaze fixations by using an eye-tracker (Tobii Pro Glasses 2) during the whole experimental  
553 session. The sampling rate was 50Hz.

554

555

556 **Results and Discussion**557 *Accuracy*558 **Figure 3**559 *Mean Accuracy and Individual Plots in Studies 2A and 2B*

560

561 *Note.* Panel (A): Recognition trials in Study 2A. Panel (B): Recall trials in Study 2A. Panel

562 (C) Recall trials in Study 2B. Individual dots indicate individual data. Y-axis error bars

563 represent standard error of mean. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ . \*\*\*\*  $p < .0001$ . n.s.  
564 non-significant.

565

566 Figure 3 shows the mean accuracy rates in Studies 2A and 2B (left panel: recognition  
567 trials; middle panel: recall trials in Study 2A; right panel: recall trials in Study 2B). A series  
568 of one-way repeated-measures ANOVA on each accuracy rates revealed significant main  
569 effects of cue condition on recognition accuracy in Study 2A (Panel A),  $F(2, 62) = 15.211$ ,  $p$   
570  $< .0001$ , on recall accuracy in Study 2A (Panel B),  $F(2, 62) = 20.594$ ,  $p < .0001$ , and on  
571 recall accuracy in Study 2B (Panel C),  $F(2, 62) = 33.481$ ,  $p < .0001$ .

572 A pre-registered planned comparison revealed that a single-item retro-cueing effect  
573 (single-cue vs. neutral-cues) was significant on recognition accuracy in Study 2A,  $t(31) =$   
574  $5.249$ ,  $p < .0001$ , Cohen's  $d = 1.031$ , its 95%  $CI = [0.495, 1.567]$ , on recall accuracy in Study  
575 2A,  $t(31) = 6.722$ ,  $p < .0001$ , Cohen's  $d = 1.258$ , 95%  $CI = [0.706, 1.810]$ , and on recall  
576 accuracy in Study 2B,  $t(31) = 8.020$ ,  $p < .0001$ , Cohen's  $d = 1.832$ , 95%  $CI = [1.229, 2.435]$ .  
577 Thus, a robust single-item retro-cueing effect was replicated in our study, for both  
578 experiments.

579 Regarding the multi-item retro-cueing effect (multi-cues vs. neutral-cues), a pre-  
580 registered planned comparison revealed that the accuracy difference was not significant on  
581 the recognition measurement in Study 2A,  $t(31) = 1.030$ ,  $p = .311$ , n.s., Cohen's  $d = 0.222$ ,

582 95%  $CI = [-0.281, 0.726]$ . However, its effect was significant on the recall measurement both  
583 in Study 2A,  $t(31) = 3.747$ ,  $p = .0007$ , Cohen's  $d = 0.768$ , 95%  $CI = [0.247, 1.289]$  and  
584 Study 2B,  $t(31) = 3.186$ ,  $p = .0032$ , Cohen's  $d = 0.772$ , 95%  $CI = [0.251, 1.294]$ . A pre-  
585 registered internal meta-analysis (Ueno et al., 2016) across the two studies by a random-  
586 effect model (Cumming, 2012) found that the integrated effect size was Hedges'  $g$  of 0.663,  
587 95%  $CI = [0.397, 0.928]$  on recall accuracy. Thus, whilst single-probe recognition could not  
588 detect a significant multi-item retro-cueing effect, probed recall robustly did. Our preferred  
589 interpretations regarding the differences between testing methods focus on their differing  
590 chance levels and the relative involvement of internal control. A 2-alternative (old/new)  
591 probe recognition test would be relatively noisy as the chance level is 50%. In contrast, the 9-  
592 alternative color recall has a much lower chance level (around 11%), making a guessing  
593 strategy far less effective and potentially increasing task sensitivity. Indeed, in the recognition  
594 trials, the mean accuracy minus 1SD was below 50% (chance level) both in the multi-cues  
595 condition (49.62%) and in the neutral-cues condition (47.53%), indicating that performance  
596 levels in the recognition trials were relatively close to floor. Moreover, relative to  
597 recognition, recall is likely to involve a more effortful, demanding, and internally driven  
598 retrieval process (e.g. Allen et al., 2018; Craik & McDowd, 1987; Craik et al., 1996), and so  
599 could be more influenced by selective attentional effects, though retro-cue benefits are of  
600 course observed on recognition measures (Souza & Oberauer, 2016, for a review). Related to  
601 this, small effects may be easier to be detected using more difficult tasks, in general.

602           Next, we should also note a possibility that a multi-item retro-cueing effect might be  
603 obtained only when the two cues were presented in adjacent positions. If this is the case, a  
604 more straightforward account for the effect would be chunking/grouping of the two cued  
605 items into one rather than assuming successful attention to two items. Indeed, Heuer and  
606 Schubö (2016) observed that spatial distance of the cued items can modulate the multi-item  
607 retro-cueing effect. To address this concern, we divided the double-cue trials in terms of  
608 whether the cued positions were adjacent or not. As a result, there was not a significant  
609 difference on accuracy between the adjacent trials ( $M = 49.47\%$ ,  $SD = 15.97\%$ ) and the non-  
610 adjacent trials ( $M = 44.44\%$ ,  $SD = 13.75\%$ ),  $t(31) = 1.476$ ,  $p = .150$ , *n.s.* Accuracy in each  
611 case was significantly higher than that in the neutral-cue condition ( $ps < .05$ ). Therefore,  
612 although there might be a chunking/grouping mechanism for the cued items in case of the  
613 adjacent positions, such an account cannot fully explain the higher accuracy in the double-cue  
614 condition than the neutral-cue condition.

615

#### 616 ***Monte-Carlo Simulation on Our Accuracy Pattern***

617           Finally, it is useful to establish whether the significant findings observed in our  
618 Studies 2A and 2B may also be explained in terms of a single-item focusing strategy as was

619 suggested in the Study 1 simulations. Thus, we conducted the Monte-Carlo simulation with  
620 the experimental parameters derived from our Studies 2A and 2B. As a result, the chance  
621 rates were 44% to obtain a significant multi-cueing effect when only one of the cued items  
622 was strategically focused on, an estimated rate that was equivalent to those we established for  
623 existing literature. Thus, despite the instruction to attend multiple items in terms of eye  
624 movements, the accuracy patterns could in principle be caused by single-item enhancement.  
625 We return to this issue after analyzing the gaze data in the next section.

626

## 627 *Fixation*

628 **Gaze Analysis.** We used Tobii Pro Lab Analyzer software (Tobii Pro AB, Stockholm,  
629 Sweden) for eye-tracking analysis in Study 2B, following the method of Jongerius et al.  
630 (2021). Tobii Pro Glasses 2 captured the visual environment in which each participant was  
631 looking and Tobii Pro Lab Analyzer allows analysis of this gaze information. The manual  
632 analysis on this software starts with taking a snapshot of the representative video frame that  
633 captures the areas that one wants to analyze (i.e., PC screen in our case). Then, onto this  
634 snapshot, we manually drew six regions of interests, using an area-of-interest tool. In our  
635 case, we first drew an invisible hexagon on the PC screen within the snapshot, such that  
636 tangents of the six circles (i.e., six targets or six cues) were the mid-points of the six sides of

637 the invisible hexagon. Then, this hexagon was divided into six isosceles triangles whose three  
638 apexes were the center of the invisible hexagon (i.e., center screen) and the apexes of the  
639 hexagon, respectively. The resultant six isosceles triangles were the six areas of interest and  
640 are shown in the top-row of Figure 4.

641         Once the areas of interests are drawn on the snapshot, then one can map eye gaze data  
642 during the time of interests (in our case, 1,000ms between the offset of the cues and the onset  
643 of a probe) onto the snapshot by using the Automatic Mapping Function. Then, various gaze  
644 information can be estimated in each time window per an area of interest. Since we aimed to  
645 visualize the distribution of gazes across the cued positions, we estimated the duration of  
646 fixations in each area of interest during the time of interest. The definition of a fixation varies  
647 depending on studies, but we used the default I-VT (Velocity-Threshold identification  
648 fixation filter) of Tobii Pro Lab Analyzer (Komogortsev et al., 2010; Salvucci & Goldberg,  
649 2000). Specifically, the I-VT fixation classifier applies an angular velocity threshold (30  
650 degrees/second) on each data point. Data points with angular velocity below the threshold  
651 value were classified as being part of a fixation.

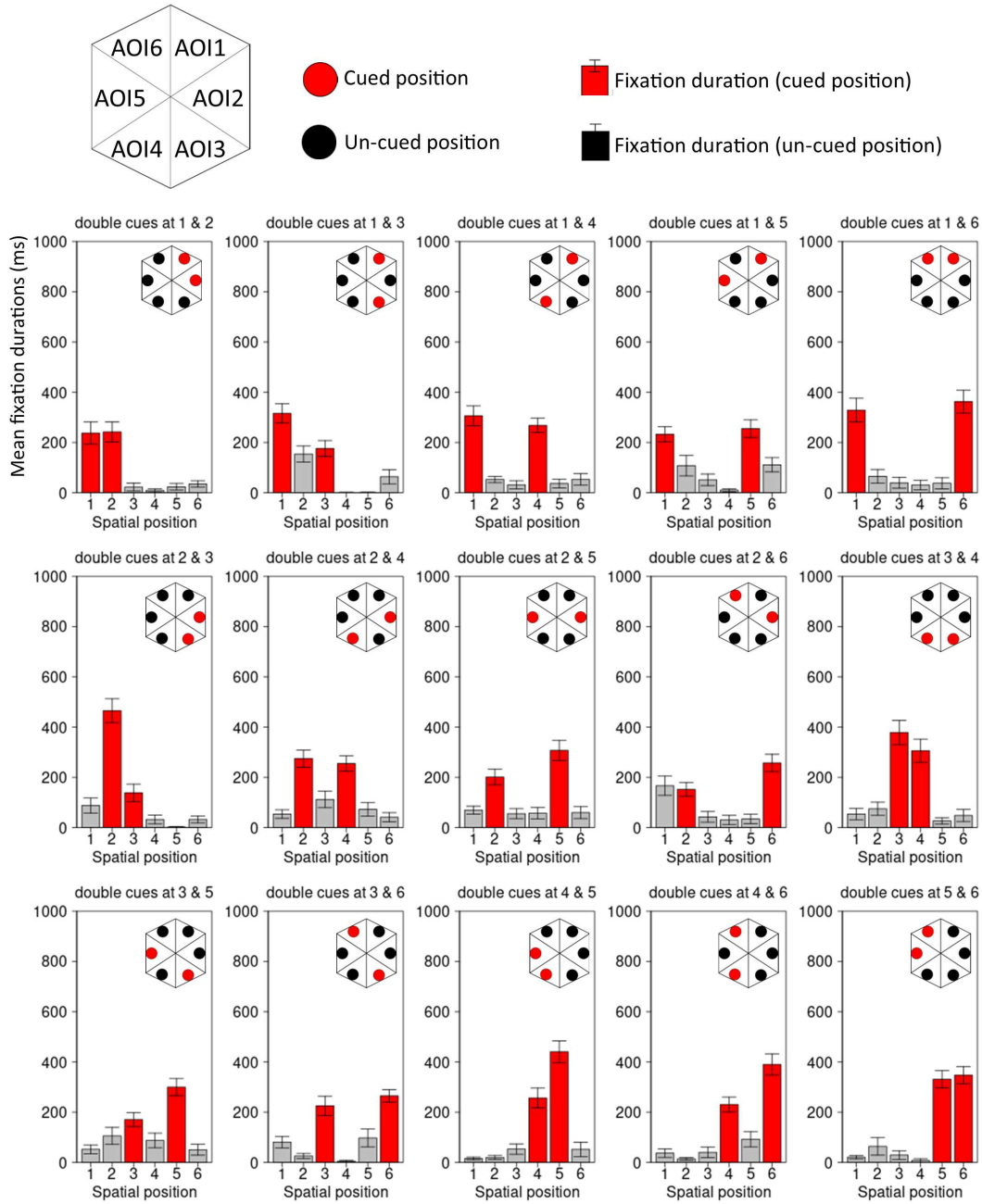
652

653

654 **Figure 4**

655 *Areas of Interests (top) and Fixation Durations in Each Area in the Double Retro-Cues*

656 *Condition.*



657

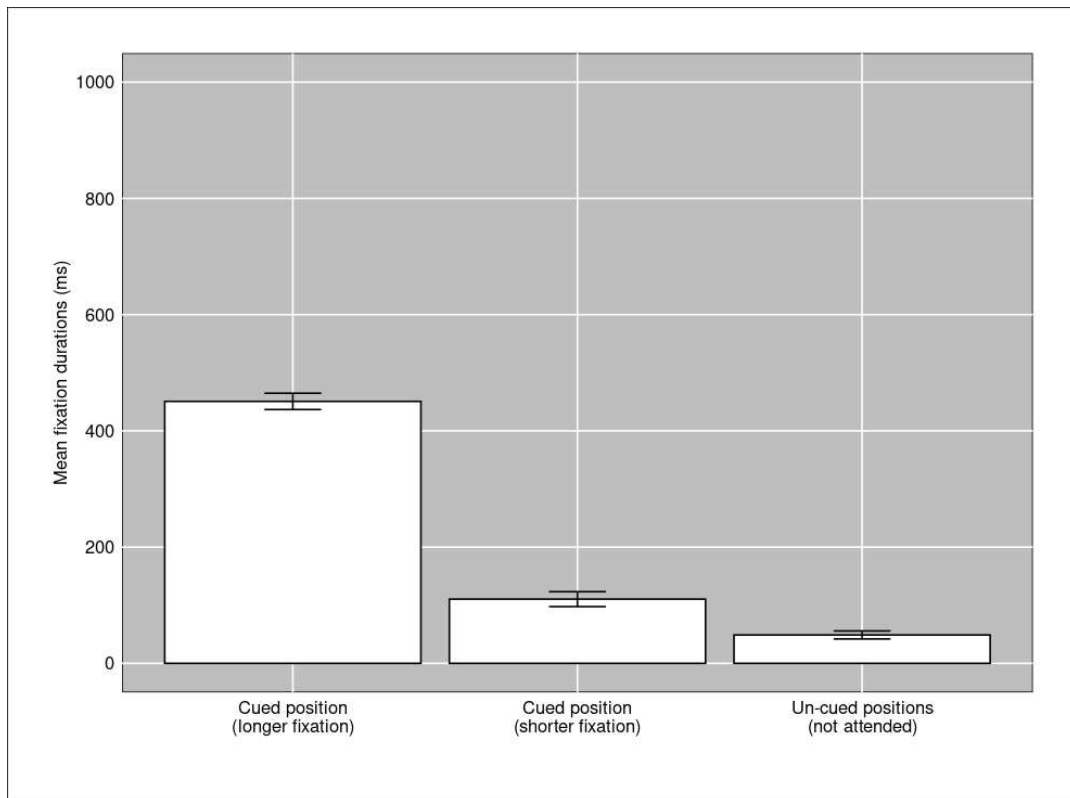
658

659 **Outcome.** Figure 4 shows the areas of interests where fixation durations were analyzed (top  
660 row), and the duration of fixations in each area after double-cues were presented. The figure  
661 for the single-cueing condition is available online (DOI 10.17605/OSF.IO/DRBXT). The  
662 fixation data were averaged over the trials per participant and then averaged over participants.  
663 The bars and the circles are shown in red when a cue had appeared at that spatial position.  
664 This clearly shows that the mean fixation values in the multi-cued positions (red) were higher  
665 than the un-cued positions. However, to echo the focus of the Monte Carlo simulation in  
666 Study 1, this *averaged* data pattern over the trials might emerge even if a participant focused  
667 on only one of the multiple cued positions within a trial. For example, suppose positions 1  
668 and 2 were cued, and then a participant's gaze was only fixated to position 1 in trial  $n$  and  
669 position 2 in trial  $n+1$ . Averaging these two trials would still result in longer fixation  
670 durations for positions 1 and 2, relative to other positions, again potentially producing  
671 artificial evidence for multi-cueing. It is important to show that each of the two cued  
672 positions received longer fixation durations than the uncued positions within a trial. Thus, we  
673 sorted the fixation data on the two cued positions within a multi-cued trial, and categorized  
674 the data as follows: (1) the longer fixation duration among the two cued positions, (2) the  
675 shorter fixation duration among the two cued positions, and (3) the fixation duration for the  
676 un-cued positions. Figure 5 shows the mean of these sorted and categorized fixation  
677 durations, averaged across trials and participants. Interestingly, the mean gaze fixation

678 durations on the two cued positions were typically not equivalent (i.e., 450.74 ms for the  
679 longer fixation and 110.51 ms for the shorter fixation). Nevertheless, even the shorter-fixated  
680 cued position received a longer duration of gaze fixations than the uncued positions  
681 (48.83ms),  $t(31) = 3.280$ ,  $p = .003$ , Cohen's  $d = 1.065$ ,  $95\%CI = [0.526, 1.603]$ . Therefore,  
682 each of the cued positions received longer fixation durations than the un-cued positions,  
683 indicating participants followed the instruction to attend both cued positions. In other words,  
684 in contrast to the prior literature, we can confidently argue that participants deliberately  
685 avoided a single-item focusing strategy, at least in terms of overt (i.e., gaze-based) attention.  
686 We discuss this issue further in the general discussion in conjunction with the simulation  
687 outcomes for our accuracy data.

688

689

690 **Figure 5**691 *Mean Fixation durations on the two cued positions (sorted by fixation length within a trial)*692 *and those on the uncued positions.*

693

694 *Note. Y-axis error bars represent standard errors of means*

695

696 The imbalanced fixation durations for the cued items in Figure 5 indicate that

697 although multiple cued items were being attended, rather than allocating equal attention to

698 each cued item, a spontaneous form of prioritization may have been applied to one of the two

699 cues. This might be part of the reason why the literature has difficulty in finding reliable

700 multi-item retro-cueing effect, particularly when retro-cues were presented simultaneously

701 (e.g., Makovski & Jiang, 2007; see also Delvenne & Holt, 2012; Heuer & Schubö, 2016;  
702 Matsukura & Vecera, 2015, for a significant multi-item retro-cueing effect only when a  
703 specific boundary condition was satisfied).

704

### 705 *Further Analyses on the Imbalanced Fixation Durations between the Cued Positions*

706         The imbalanced fixation durations between the two cued positions motivated us to  
707 further investigate the differences between the cued positions. Thus, although not pre-  
708 registered, we conducted the following exploratory analyses. The first involves the order of  
709 gaze visits. Specifically, we investigated whether the first-fixated cued item or the second  
710 item tend to receive the longer fixation. The left panel of Figure 6 shows the total gaze  
711 fixation durations for the multi-cued spatial positions, divided in terms of whether each cued  
712 position was first-fixated or second-fixated. As a result, the first-fixated cued position  
713 received approximately 410ms of fixation durations in total (averaged over the trials and over  
714 the participants) whilst the second-fixated cued item received approximately 140ms of  
715 fixation durations. Thus, participants spent longer looking at the first-fixated position.  
716 Related to this, the right panel of Figure 6 also shows the split data in terms of the order of  
717 the gaze fixation visit, but this one shows only the fixation durations during the initial visit to  
718 each position. Extracting the fixation durations only for the initial visit time is informative as

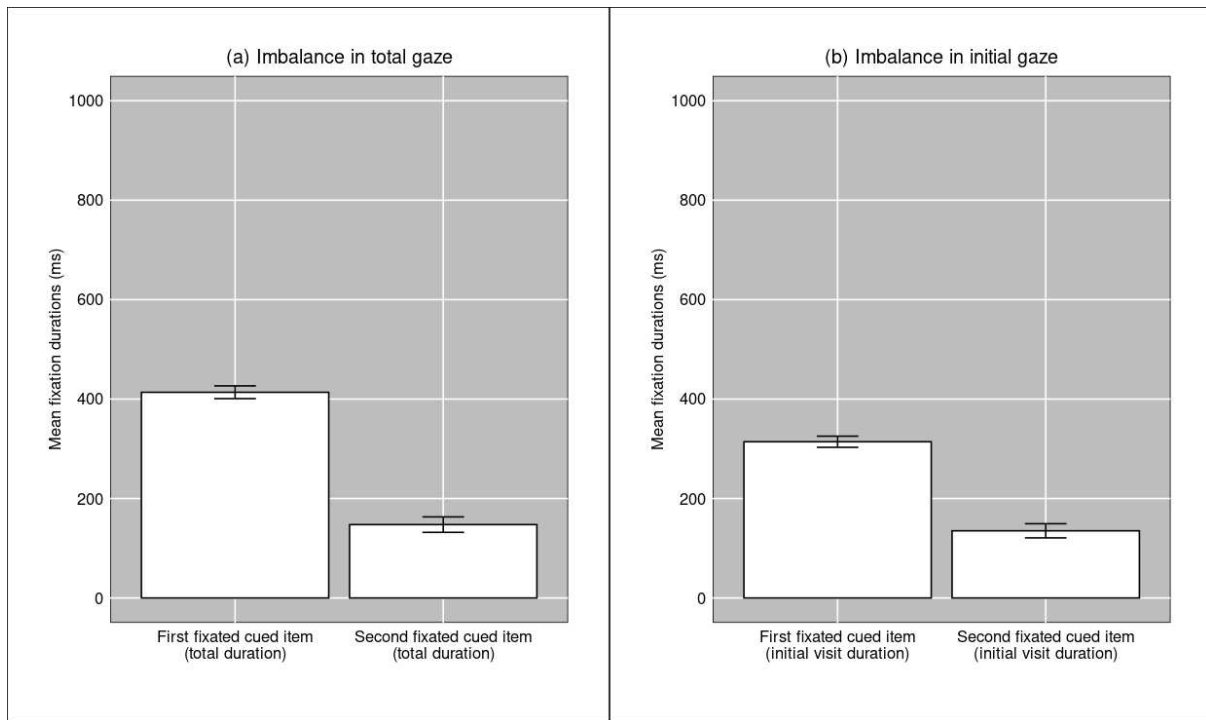
719 some participants may visit the same position multiple times during the retention interval. As  
720 a result, the gaze fixation duration in the initial visit was approximately 300ms on average in  
721 the first-fixated cued position whilst 135ms on average for the second-fixated cued position.  
722 In other words, the difference between the left and the right panels of Figure 6 indicates the  
723 gaze fixation duration at the second (and subsequent) visit to each position (i.e.,  
724 approximately 110ms on average in the first-fixated cued position whilst only 5ms in the  
725 second-fixated cued position). Taken together, we can characterize gaze patterns as follows;  
726 the first-fixated position received a longer duration of gaze fixations (300ms) at the initial  
727 visit, after which the second one received a relatively shorter duration of gaze fixations  
728 (135ms on average), after which gaze briefly returned to the first position (110ms on average  
729 – i.e., the difference between the left and the right halves of Figure 6). Thus, overt gaze-based  
730 retrospective attention seems to be allocated to multiple positions sequentially.

731

732

733 **Figure 6**

734 *Mean fixation durations to the two cued positions (divided in terms of the order of fixation*  
 735 *visit within a trial)*



736

737 *Note. Panel A (left) shows the total fixation duration in each trial; Panel B (right) shows*  
 738 *fixation duration at the time of initial visit to each position. Y-axis error bars represent*  
 739 *standard error of mean.*

740

741 A tempting idea from these outcomes is that the first-fixated item, whose position  
 742 received longer fixation durations, may show higher accuracy when it is probed. We tested  
 743 this in two ways. First, we split the accuracy data in the multi-cues condition in terms of

744 whether the probed position was first-fixated (*Mean* accuracy = 46.90%, *SD* =13.89%) or  
745 second-fixated (*Mean* = 45.47%, *SD* =14.12%). A paired *t*-test did not detect a significant  
746 difference,  $t(31) = 0.514, p = .611, d = 0.102, 95\%CI = [-0.40, 0.60]$ . Secondly, Figure 7  
747 plots the relationship between accuracy in each trial (a binary measure) and the total fixation  
748 duration on the probed position, including only the cued and probed position. To examine the  
749 effect of fixation duration, we conducted generalized linear mixed-effect modelling. The  
750 fixed effect was the fixation duration; the response variable was accuracy; the random  
751 variables were participants (random-intercept and random-slope) and the probed position  
752 (random-intercept). As a result, fixation duration did not significantly predict the log-odds of  
753 accuracy: coefficient = -0.0003,  $z$ -value = -1.438,  $p = .150$ . The black curve in Figure 7  
754 represents the logistic curve (fixed-effect only) estimated by the GLMM, appearing as a  
755 straight line due to the non-significant effect of the predictor. The non-significant effect of  
756 predictor was found even when the data were split and analyzed by probed spatial position  
757 (available online: DOI 10.17605/OSF.IO/DRBXT). Taken together, our data suggests that  
758 fixation duration did not predict accuracy for cued items, and relatedly, the mean accuracy in  
759 the first-fixated position, which received longer fixation durations, was not significantly  
760 different from that in the second-fixated position.

761           Importantly, the equivalent accuracy rates between the first-fixated position and the  
762 second-fixated position despite their imbalanced fixation durations can help inform regarding  
763 the underlining processes in multi-cueing effect. As mentioned in the introduction of Study 2,

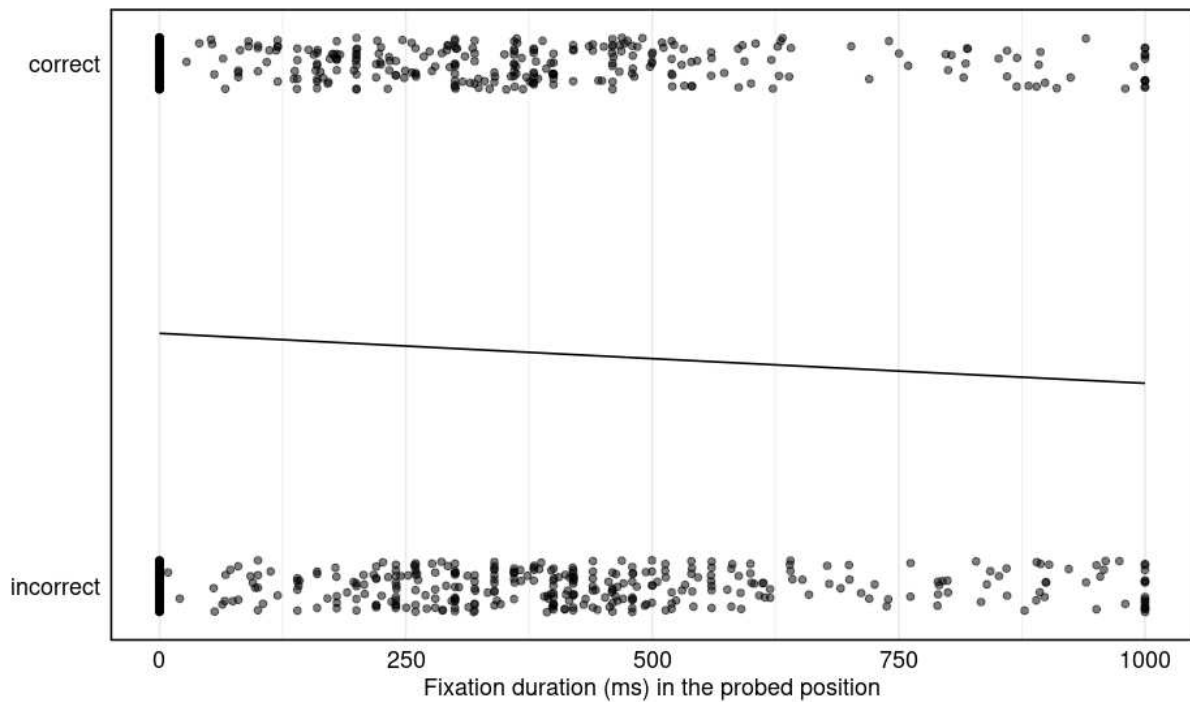
764 fixating gazes at both positions demonstrates that participants made effort to avoid a single-  
765 item focusing strategy, but this does not necessarily mean that both items benefited from  
766 gaze-based attention. Instead, only one of them might benefit. However, if this were the case,  
767 it would be expected that the item in the longer-fixated position would be enhanced, which  
768 was not the case. Instead, our interpretation is that both items were enhanced by the  
769 sequentially allocation of gaze-based attention. Rather than gaze order or duration, gaze  
770 visitation towards each position determines the enhancement of memory accuracy.

771

772

773 **Figure 7**

774 *Trial-by-trial scatterplots for the relationship between fixation duration on the probed*  
 775 *position and its accuracy.*



776

777 *Note. Each point was adjusted with a small amount of random jitter (along the y-axis) to*  
 778 *reduce overlap. The line represents the logistic regression curve (see main text).*

779

780

### General Discussion

781 Several studies have investigated the multi-item retro-cueing effect on visual working  
 782 memory. However, after Makovski and Jiang (2007)'s initial study, almost all the studies

783 have speculated on the reasons for the inconsistent effects that have observed, and some have  
784 attributed them to methodological variations between studies. We hypothesized that a single-  
785 item focusing strategy might be one reason for the inconsistencies and aimed to estimate the  
786 chance rates to obtain a statistically significant multi-item retro-cueing effect despite the use  
787 of a single-item strategy. Study 1 employed a Monte-Carlo simulation and revealed that the  
788 estimated chance rates were much higher than the conventional alpha level (5%) under the  
789 experimental settings of the existing literature. Such a high rate is consistent with the  
790 presence of mixed outcomes in the literature. The estimated chance rates represent the rate of  
791 possible misinterpretation if statistical significance is interpreted as evidence for  
792 retrospectively attending multiple items. Studies 2A and 2B investigated the multi-item retro-  
793 cueing effect in human participants, but this time a single-item focusing strategy was  
794 discouraged by instructing distribution of gaze fixations across all the cued positions. Eye-  
795 tracking in Study 2B confirmed that participants followed this instruction in each trial (see  
796 Figure 5). Thus, we collected plausible evidence that participants at least overtly attended to  
797 multiple items in each trial. Moreover, although a 2-alternative recognition test could not find  
798 a significant multi-item retro-cueing effect (indicating a relative lack of sensitivity), this  
799 effect was found in cued recall across the two studies.

800 Overall, our simulation indicates that prior evidence for multi-item cueing might often  
801 reflect a single item strategy. Possibly stronger evidence for a significant multi-item retro-  
802 cueing effect can be found using cued recall and the explicit direction to fixate on all cued

803 locations. Eye-tracking confirms this, while also uncovering more complex patterns of gaze  
804 behaviour that do not directly map onto memory accuracy. Our study therefore at least  
805 somewhat reduces the possibility for a single-item focusing strategy account to explain the  
806 multi-item retro-cueing effects. However, these findings come with the caveat that simulation  
807 of our own accuracy data also indicated a relatively high chance rate for the apparent multi-  
808 cue effect on accuracy to in fact reflect a covert single item attentional focus within working  
809 memory. We will return to this possibility at the end of the discussion.

810         Among counterarguments or concerns regarding the explanatory power of a single-  
811 item focusing strategy, two to consider are cue validity level and possible interactions with  
812 other factors. First, is a single-item focusing strategy effective even when cue validity level is  
813 low? Our Monte-Carlo simulation demonstrated that as far as the cue-valid trials are scored,  
814 the chance rates to obtain a significant multi-item retro-cueing effect by this strategy was still  
815 high even in the lower cue validity level (e.g., Li & Saiki, 2014). Secondly, one may be  
816 tempted to exclude this strategy account if the multi-item retro-cueing effect is modulated by  
817 another factor (i.e., is involved in an interaction), such as the spatial positions/distances of the  
818 cues (e.g., Heuer & Schubö, 2016). However, we demonstrated a high chance rate for a  
819 single-item focusing strategy to result in a significant multi-item retro-cueing effect in one  
820 condition alongside a non-significant effect in the other condition (i.e., to produce an  
821 interaction). Therefore, one should be cautious about interpreting multi-item retro-cueing  
822 effects even if these effects appear to be modulated by another factor.

823           When developing optimal measures to detect, analyze, and interpret multi-item retro-  
824 cueing effects, what factors do we need to consider? First, task instructions should discourage  
825 the use of a single-item focusing strategy as much as possible, and one should obtain as much  
826 information as possible regarding the observance of the instruction, including gaze fixations.  
827 Another way might be to use a value-based prioritization instruction (e.g., Allen & Ueno,  
828 2018; Atkinson et al., 2018; Hu et al., 2014; Allen et al., 2024, for a review) rather than  
829 cueing. Value-based prioritization may represent a lenient form of selective attention  
830 (Oberauer, 2019), and participants are assumed not to completely neglect the unprioritized  
831 items (Allen et al., 2024). For example, Allen and Atkinson (2021) found some evidence for  
832 retrospectively applied prioritization of the most recently encountered item in a sequence,  
833 though note that retrospective value effects may be somewhat smaller in magnitude than  
834 those of predictive retro-cueing (Hautekiet et al., 2024). Investigating possible multi-item  
835 retro-prioritization effects (i.e., two items receiving a higher reward) could be a useful future  
836 development.

837           A second factor to bear in mind is the size of the single-cueing effect (i.e., single-cue  
838 condition vs. neutral-cues condition). The larger this effect is, the stronger the explanatory  
839 power of a strategy-based account is (see Table 1 and Figure 1). Interestingly, the data from a  
840 recently published article (DiPuma et al., 2023) seems to be consistent with this relationship.  
841 Specifically, three experiments in DiPuma et al. (2023) could not find a significant multi-item  
842 retro-cueing effect on color/orientation memory accuracy (i.e., the absolute angular

843 difference). However, if we look at the precision measurement (i.e., how precisely one can  
844 recall the feature of the probed stimulus at a given trial: Bays et al., 2009), then it seems to be  
845 higher in the multi-cue condition than the neutral-cue condition in Experiments 1 and 3, but  
846 such a tendency was not observed in Experiment 2. Interestingly, the effect sizes of the  
847 single-cueing effect were large in their Experiments 1 and 3 but small in Experiment 2.

848         We observed that overt gaze-based attention was allocated to multiple cued positions.  
849 Additionally, our exploratory analyses revealed that gaze fixation durations were imbalanced  
850 between the two cued positions and that a first-fixated position received longer fixation  
851 durations at the time of initial visit than the second-fixated item. Moreover, eye gaze  
852 appeared to return to the first-fixated position, after the second-fixating position. Thus, overt  
853 gaze-based retrospective attention is not allocated simultaneously, and is instead allocated  
854 sequentially to each of the cued items. Furthermore, fixation duration did not predict memory  
855 accuracy. Our interpretation is that gaze visit determines the cue-based enhancement of  
856 accuracy, rather than gaze duration. Once gaze is fixated to each position, refreshing  
857 processing (Souza et al., 2015) for that item starts. We note here that the non-significant  
858 fixation-accuracy relationship is inconsistent with Loaiza and Souza (2022, Study 1B), who  
859 found a positive relationship between the fixation durations and the accuracy when eye gaze  
860 to the sequentially-presented cued positions was explicitly instructed, though they presented  
861 multi-cues sequentially whilst our study presented them simultaneously. Moreover, Loaiza  
862 and Souza used a continuous scale for reporting the color of a memorized item whilst we

863 used a 9-alternative recall measure. These methodological differences may help explain the  
864 discrepancies in findings.

865         What other implications might we draw regarding possible mechanisms underlying  
866 retro-cueing effects? First, Makovski and Jiang (2007) discussed that a central bottleneck  
867 operates in consolidation of visual working memory, such that only one item at a time is  
868 processed and the consolidation process interferes with the maintenance of other items  
869 (Griffin & Nobre, 2003). We found a clear difference in accuracy between the single-cue  
870 condition and the double cue condition. Thus, there may indeed be some forms of mutual  
871 interference (or cost) to retrospectively attend multiple items (though it should be noted that  
872 this comparison is confounded by probe validity and was not a primary focus of the current  
873 work). Secondly, other studies provide more complex theories regarding working memory  
874 and attention by providing boundary conditions for the multi-item retro-cueing effects  
875 (Delvenne & Holt, 2012; Heuer & Schubö, 2016; Matsukura & Vecera, 2015), but such  
876 complexity appears to be unnecessary. Moreover, some studies have found positive effects of  
877 sequentially presented multiple retro-cues (Li & Saiki, 2014; Souza et al., 2015, but see Van  
878 Moorselaar et al., 2015). Therefore, one might be tempted to speculate that different  
879 principles apply depending on the presentation format of the cues. However, once again, such  
880 complexity is likely unnecessary. Thus, an account for single-item retro-cueing effects may  
881 be extended to multi-cueing without additional complexity. For example, Souza and  
882 Oberauer (2016) identified the following four accounts to explain a retro-cueing effect:

883 retrospectively attended representations are strengthened; un-attended items are removed  
884 from working memory; a head-start retrieval is provided in the accumulation of evidence for  
885 the attended items; and attended items are more protected from interference. Although our  
886 studies cannot differentiate between these accounts, we would argue that they can potentially  
887 be extended to explain multi-item retro-cueing effects without additional assumptions.

888         However, despite the trial-based eye-tracking evidence illustrating direction of spatial  
889 attention toward the locations of multiple cued items, simulation indicates that the overall  
890 accuracy patterns, which were aggregated across trials and participants, could in principle be  
891 produced by single item enhancement on each trial. This is because we cannot argue with  
892 100% confidence that overt eye-movement necessarily translates into genuine multi-item  
893 effects in working memory in a way that has measurable benefits on recall performance (see  
894 also Figure 7). Thus, other forms of single-item focusing strategy rather than gaze-fixation  
895 might have contributed to aggregated accuracy pattern. For example, one could argue that  
896 participants looked at both positions because the experimenter instructed them to do so, but  
897 that directed gaze did not play much of a role in enhancing accuracy. Instead, enhancements  
898 in mean accuracy over trials in the multi-cue condition were driven by a covert single item  
899 focusing strategy. Under this account, although participants clearly observed the instruction  
900 to allocate overt gaze-based attention to multiple items, covert attentional processing within  
901 working memory did not follow on from this. Such an account is inconsistent with findings in  
902 the literature that have demonstrated an active and facilitatory role of gaze position in

903 memory after offset of studied items (Ferreira et al., 2008; Johansson & Johansson, 2014;  
904 Laeng et al., 2014; van Ede et al., 2019, see also Loaiza & Souza, 2022, for the effect of  
905 instructional gaze), but we cannot completely reject it as a possibility. This shows that, even  
906 after adjusting task context and measurement methods to increase the probability of multi-cue  
907 implementation, it remains challenging to clearly adjudicate between single and multiple item  
908 interpretations. Thus, we see this as a call for the field to develop more sophisticated and  
909 appropriate methods to derive more confident conclusions regarding genuine multiplicity in  
910 working memory. Though not providing a perfect solution, our explicit instruction and gaze-  
911 monitoring approach forms the first step to disentangle at least an overt gaze-based allocation  
912 of attention from single-item focusing strategies. Future studies should aim to specify and de-  
913 confound other forms of single-focus strategies to isolate genuine multi-cueing effects.

914

## 915 **Conclusions**

916       There is mixed evidence in the literature regarding whether multiple retro-cues can  
917 facilitate working memory for each of the cued items. Our Monte-Carlo simulation indicated  
918 that a potentially misleading multi-cueing effect might emerge using a single item focus  
919 strategy, which could go some way to accounting for the inconsistent evidence to date. Two  
920 experiments then demonstrated that a multi-item cueing effect in aggregated accuracy data

921 could indeed be observed with explicit instruction to direct gaze to all cued locations. Eye-  
922 tracking confirmed engagement with this instruction. However, simulation suggests that our  
923 observed accuracy patterns could in principle be produced by single-item enhancement on  
924 each trial. Thus, future studies should aim to develop improved methods to disentangle  
925 possible multi-item cueing from other forms of single item focusing strategies. Nevertheless,  
926 with explicit instruction and gaze monitoring, the current study forms the first step to isolate  
927 genuine retro-cueing multiplicity in working memory.

928

929

930

**References**

931

932 Allen, R. J., & Atkinson, A. L. (2021). Retrospective and Prospective prioritization in visual  
933 working memory. *PsrArXiv*. <https://doi.org/10.31234/osf.io/4x8zu>

934 Allen, R. J., Atkinson, A., & Hitch, G. J. (2024). Getting value out of working memory  
935 through strategic prioritisation; implications for storage and control. *Quarterly Journal*  
936 *of Experimental Psychology*. <https://doi.org/10.1177/17470218241258102>

937 Allen, R. J., & Ueno, T. (2018). Multiple high-reward items can be prioritized in working  
938 memory but with greater vulnerability to interference. *Attention, Perception &*  
939 *Psychophysics*. <https://doi.org/10.3758/s13414-018-1543-6>

940 Atkinson, A. L., Berry, E., Waterman, A. H., Baddeley, A. D., Hitch, G. J., & Allen, R. J.  
941 (2018). Are there multiple ways to direct attention in visual working memory? *Annals of*  
942 *the New York Academy of Sciences*. <https://doi.org/10.1111/nyas.13634>

943 Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2018). Exploring the sentence advantage in  
944 working memory: Insights from serial recall and recognition. *Quarterly Journal of*  
945 *Experimental Psychology*, 71(12), 2571-2585.

946 <https://doi.org/10.1177/1747021817746929>

- 947 Atkinson, A. L., Oberauer, K., Allen, R. J., & Souza, A. S. (2022). Why does the probe value  
948 effect emerge in working memory? Examining the biased attentional refreshing  
949 account. *Psychonomic Bulletin & Review*, 29(3), 891-900.  
950 <https://doi.org/10.3758/s13423-022-02056-6>
- 951 Barras, C., & Kerzel, D. (2017). Salient-but-irrelevant stimuli cause attentional capture in  
952 difficult, but attentional suppression in easy visual search. *Psychophysiology*, 54(12).  
953 <https://doi.org/10.1111/psyp.12962>
- 954 Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working  
955 memory is set by allocation of a shared resource. *Journal of Vision*, 9(10).  
956 <https://doi.org/10.1167/9.10.7>
- 957 Craik, F.I.M., Govoni, R., Naveh-Benjamin, M., & Anderson, N.D. (1996). The effects of  
958 divided attention on encoding and retrieval processes in human memory. *Journal of*  
959 *Experimental Psychology: General*, 125, 159-180.
- 960 Craik, F. I., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of*  
961 *Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 474.
- 962 Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and*  
963 *meta-analysis*. Routledge.

- 964 Delvenne, J. F., & Holt, J. L. (2012). Splitting attention across the two visual fields in visual  
965 short-term memory. *Cognition*, *122*(2), 258–263.  
966 <https://doi.org/10.1016/j.cognition.2011.10.015>
- 967 DiPuma, A., Lockhart, H. A., Emrich, S. M., & Ester, E. F. (2023). Retrospective cue  
968 benefits in visual working memory are limited to a single location at a time. *Attention,*  
969 *Perception, and Psychophysics*, *85*(5). <https://doi.org/10.3758/s13414-023-02661-0>
- 970 Griffin, I. C., & Nobre, A. C. (2003). Orienting Attention to Locations in Internal  
971 Representations. *Journal of Cognitive Neuroscience*, *15*(8), 1176–1194.  
972 <https://doi.org/10.1162/089892903322598139>
- 973 Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing.  
974 *Trends in Cognitive Sciences*, *12*(11). <https://doi.org/10.1016/j.tics.2008.07.007>
- 975 Hautekiet, C., Langerock, N., & Vergauwe, E. (2024). Prioritization in Visual Working  
976 Memory: An Investigation of Distractor Susceptibility and Different Prioritization  
977 Modes. *Open Science Framework*.
- 978 Heuer, A., & Schubö, A. (2016). Feature-based and spatial attentional selection in visual  
979 working memory. *Memory & Cognition*, *44*(4), 621–632.  
980 <https://doi.org/10.3758/s13421-015-0584-5>

- 981 Hu, Y., Hitch, G. J., Baddeley, A. D., Zhang, M., & Allen, R. J. (2014). Executive and  
982 perceptual attention play different roles in visual working memory: Evidence from  
983 suffix and strategy effects. *Journal of Experimental Psychology: Human Perception and*  
984 *Performance*, 40(4), 1665–1678. <https://doi.org/10.1037/a0037163>
- 985 Johansson, R., & Johansson, M. (2014). Look Here, Eye Movements Play a Functional Role  
986 in Memory Retrieval. *Psychological Science*, 25(1).  
987 <https://doi.org/10.1177/0956797613498260>
- 988 Jongerius, C., Callemein, T., Goedemé, T., Van Beeck, K., Romijn, J. A., Smets, E. M. A., &  
989 Hillen, M. A. (2021). Eye-tracking glasses in face-to-face interactions: Manual versus  
990 automated assessment of areas-of-interest. *Behavior Research Methods*, 53(5).  
991 <https://doi.org/10.3758/s13428-021-01544-2>
- 992 Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010).  
993 Standardization of automated analyses of oculomotor fixation and saccadic behaviors.  
994 *IEEE Transactions on Biomedical Engineering*, 57(11).  
995 <https://doi.org/10.1109/TBME.2010.2057429>
- 996 Laeng, B., Bloem, I. M., D’Ascenzo, S., & Tommasi, L. (2014). Scrutinizing visual images:  
997 The role of gaze in mental imagery and memory. *Cognition*, 131(2).  
998 <https://doi.org/10.1016/j.cognition.2014.01.003>

- 999 Li, Q., & Saiki, J. (2014). The effects of sequential attention shifts within visual working  
1000 memory. *Frontiers in Psychology*, 5(AUG). <https://doi.org/10.3389/fpsyg.2014.00965>
- 1001 Loaiza, V. M., & Souza, A. S. (2022). The eyes don't have it: Eye movements are unlikely to  
1002 reflect refreshing in working memory. *PLoS ONE*, 17(7 July).  
1003 <https://doi.org/10.1371/journal.pone.0271116>
- 1004 Makovski, T., & Jiang, Y. V. (2007). Distributing versus focusing attention in visual short-  
1005 term memory. *Psychonomic Bulletin & Review*, 14(6), 1072–1078.  
1006 <https://doi.org/10.3758/BF03193093>
- 1007 Matsukura, M., & Vecera, S. P. (2015). Selection of multiple cued items is possible during  
1008 visual short-term memory maintenance. *Attention, Perception & Psychophysics*, 77(5),  
1009 1625–1646. <https://doi.org/10.3758/s13414-015-0836-2>
- 1010 Oberauer, K. (2019). Working memory and attention - A conceptual analysis and review. In  
1011 *Journal of Cognition* (Vol. 2, Issue 1). <https://doi.org/10.5334/joc.58>
- 1012 Oberauer, K., & Hein, L. (2012). Attention to Information in Working Memory. *Current*  
1013 *Directions in Psychological Science*, 21(3), 164–169.  
1014 <https://doi.org/10.1177/0963721412444727>
- 1015 Ort, E., & Olivers, C. N. L. (2020). The capacity of multiple-target search. *Visual Cognition*,  
1016 28(5–8). <https://doi.org/10.1080/13506285.2020.1772430>

- 1017 Perugini, M., Gallucci, M., & Costantini, G. (2018). A Practical Primer To Power Analysis  
1018 for Simple Experimental Designs. *International Review of Social Psychology*, 31(1): 20,  
1019 1–23, DOI: <https://doi.org/10.5334/irsp.181>
- 1020 R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. R  
1021 Foundation for Statistical Computing, Vienna, Austria.
- 1022 R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R  
1023 Foundation for Statistical Computing, Vienna, Austria.
- 1024 Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking  
1025 protocols. *Proceedings of the Eye Tracking Research and Applications Symposium*  
1026 *2000*. <https://doi.org/10.1145/355017.355028>
- 1027 Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory:  
1028 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, 78(7), 1839–  
1029 1860. <https://doi.org/10.3758/s13414-016-1108-5>
- 1030 Souza, A. S., Rerko, L., & Oberauer, K. (2015). Refreshing memory traces: Thinking of an  
1031 item improves retrieval from visual working memory. *Annals of the New York Academy*  
1032 *of Sciences*, 1339(1). <https://doi.org/10.1111/nyas.12603>

- 1033 Ueno, T., Fastrich, G. M., & Murayama, K. (2016). Meta-analysis to integrate effect sizes  
1034 within an article: Possible misuse and Type I error inflation. *Journal of Experimental*  
1035 *Psychology: General, 145*(5), 643–654. <https://doi.org/10.1037/xge0000159>
- 1036 van Ede, F., Chekroud, S. R., & Nobre, A. C. (2019). Human gaze tracks attentional focusing  
1037 in memorized visual space. In *Nature Human Behaviour* (Vol. 3, Issue 5).  
1038 <https://doi.org/10.1038/s41562-019-0549-y>
- 1039 Van Moorselaar, D., Olivers, C. N. L., Theeuwes, J., Lamme, V. A. F., & Sligte, I. G. (2015).  
1040 Forgotten but not gone: Retro-cue costs and benefits in a double-cueing paradigm  
1041 suggest multiple states in visual short-term memory. *Journal of Experimental*  
1042 *Psychology: Learning Memory and Cognition, 41*(6).  
1043 <https://doi.org/10.1037/xlm0000124>
- 1044