

1 **More motivated but equally good: no effect of gamification on visual working memory**
2 **performance**

3 Maria Mystakidou & Ronald van den Berg
4

5 **ABSTRACT**

6 Gamification refers to the introduction of gaming elements such as scores and leaderboards in non-
7 gaming contexts. While there is growing evidence that gamification has positive effects on
8 intrinsic motivation and engagement, it is largely unknown whether these effects translate to
9 improved cognitive performance. Here, we examine whether gamification affects performance on
10 a visual working memory (VWM) task. In Experiment 1, we gamified a standard delayed-
11 estimation task by introducing scores and a leveling system. On each trial, the subject's estimation
12 error was mapped to a score between -100 and +100 and added to their total score. Subjects started
13 at a set size of 1 and "leveled up" to the next set size each time they had accumulated 1,500 points.
14 Post-experiment questionnaire data confirmed that subjects who performed the gamified version
15 of the task were more motivated than control subjects. However, we found no difference in VWM
16 performance between these two groups, nor between below-median and above-median motivated
17 subjects. In Experiment 2, we tested for effects of trial-by-trial manipulations of motivation on
18 VWM performance, by varying the scoring function across trials. Three scoring functions were
19 used, with maxima of 7, 21, and 101 points. At the beginning of each trial, the subject was informed
20 whether the potential reward was "low", "medium", or "high". Post-questionnaire data showed
21 that subjects were more motivated on high-reward trials. However, we found no evidence for a
22 difference in performance between the three reward levels. Our results suggest that gamification
23 increases people's motivation to carry out visual working memory tasks, but it does not necessarily
24 increase their performance.
25
26

27 **Keywords:** visual working memory; gamification; motivation; resource rationality; working
28 memory capacity; cognitive constraints
29

30 1. INTRODUCTION

31 There has recently been a surge in research on effects of “gamification”, i.e., the use of
32 gaming elements in non-gaming contexts such as learning, education, and marketing (e.g.,
33 (Deterding, Dixon, Khaled, & Nacke, 2011; Dicheva, Dichev, Agre, & Angelova, 2015; Hamari
34 & Lehdonvirta, 2010; Hanus & Fox, 2015; Subhash & Cudney, 2018)). Empirical work has
35 suggested that gamification has positive effects on a variety of psychological outcomes, such as
36 intrinsic motivation, enjoyment, engagement, and perceived competence (see Hamari et al., 2014
37 for a review). It is largely unknown, however, whether these effects are accompanied by improved
38 cognitive performance. In the present study, we examine whether gamification improves people’s
39 performance on a visual working memory (VWM) task.

40 A prerequisite for finding effects of gamification on VWM performance is that allocation of
41 VWM resources must be flexible – if it is fixed, no experimental manipulation can increase or
42 decrease VWM performance. While there has been extensive research on describing VWM
43 limitations (Brady, Konkle, & Alvarez, 2011; Luck & Vogel, 2013; Ma, Husain, & Bays, 2014),
44 few studies have asked why there are limitations in the first place. One possible answer to this
45 question is that the sustained energy that is required to keep a memory alive (Fuster & Alexander,
46 1971) induces a metabolic cost (Attwell & Laughlin, 2001; Laughlin, 2001; Sterling & Laughlin,
47 2015). In the presence of such a cost, a rational system would use its resources sparingly: resources
48 should only be invested insofar the induced cost is compensated for by expected task reward. We
49 recently formalized this idea in a resource-rational theory of VWM (Van den Berg & Ma, 2018)
50 and found that a model derived from this theory accounts well for earlier documented effects of
51 set size (Bays, Catalao, & Husain, 2009; Van den Berg, Shin, Chou, George, & Ma, 2012; Wilken
52 & Ma, 2004) and item importance (Bays, 2014; Emrich, Lockhart, & Al-Aidroos, 2017) on
53 encoding precision. Inspired by these findings, we proposed that resource-rationality may be a
54 general theory of how the brain allocates VWM resources and that VWM “limitations” are the
55 result of a cost-benefit trade-off rather than a hardwired constraint on capacity. A key implication
56 of this theory is that VWM resource allocation may be much more flexible than assumed so far.

57 If the amount of VWM resource utilized to a task is flexible, we may expect a relation
58 between a subject’s level of motivation and the amount of resource they invest in a task. Therefore,
59 we hypothesize that performance on VWM tasks can be improved by gamifying the tasks. We test
60 this hypothesis in two experiments. Experiment 1 uses a between-subject design in which we

61 compare VWM performance between subjects who perform a standard VWM task and subjects
 62 who perform a gamified version of that task. Experiment 2 uses a within-subject design, in which
 63 the number of points that a subject can earn varies across trials. To preview our results, we find
 64 that gamification increases motivation in both experiments, but we find no effect on VWM
 65 performance.

66

67 **2. EXPERIMENT 1**

68

69 **2.1. Sharing of data and analysis scripts**

70 Data and Matlab scripts related to this experiment are available at <https://osf.io/gb2kd/>.

71

72 **2.2. Participants**

73 A total of 62 participants with self-reported normal or corrected-to-normal vision were
 74 recruited using posters at various campuses of Uppsala University (Table 1). The first 40
 75 participants were randomly divided into the first three gamified groups. The remaining 22
 76 participants were recruited later and randomly divided into the control group and the last gamified
 77 group. The study was approved by the Regional Ethical Review Board in Uppsala and conducted
 78 according to the Declaration of Helsinki Principles. All participants signed informed consent and
 79 received a cinema voucher with a value of approximately \$12 for their participation.

80

81 **Table 1.** *Overview of participants. E_0 refers to the estimation error (in degrees) at which the*
 82 *scoring function mapped to a score of 0 points (see Figure 1B).*

| Condition | E_0 | Number of subjects | Age range | Mean age | Females |
|------------------------|-------|--------------------|-----------|----------|---------|
| Gamified, difficulty 1 | 60 | 13 | 20-47 | 28.0±0.5 | 7 |
| Gamified, difficulty 2 | 45 | 13 | 19-36 | 24.4±0.4 | 9 |
| Gamified, difficulty 3 | 30 | 14 | 20-43 | 26.4±0.5 | 10 |
| Gamified, difficulty 4 | 20 | 10 | 21-38 | 25.2±0.5 | 6 |
| Control | - | 12 | 20-35 | 24.6±0.4 | 9 |

83

84

85

86 **2.3. Stimuli and materials**

87 Stimuli were presented on a 23" LCD screen at a resolution of 1920×1080 pixels in a dimly
88 lighted room¹. Subjects were seated at a distance of approximately 60 cm from the screen. Two
89 centrally located concentric circles were used as a fixation point. Stimuli were dark gray ellipses,
90 presented on a light gray background (Figure 1A). The ellipses had an area of 1,000 pixels² and an
91 eccentricity (elongation) of 0.95. The stimuli were presented at a centrally located, invisible circle
92 with a radius of 220 pixels. The location of the first stimulus was drawn randomly and all other
93 stimuli were positioned such that equal spacing was ensured between any two neighboring stimuli.
94 The orientation of each stimulus was drawn from a uniform distribution on the circle, with the
95 constraint that the minimum circular distance (in degrees) between any two stimuli was at least
96 $90/N$, where N indicates the set size. This constraint was intended to discourage subjects from
97 using chunking strategies during encoding (Nassar, Helmers, & Frank, 2018). Eye movements
98 were recorded using a Tobii 4C eye tracker.

99
100

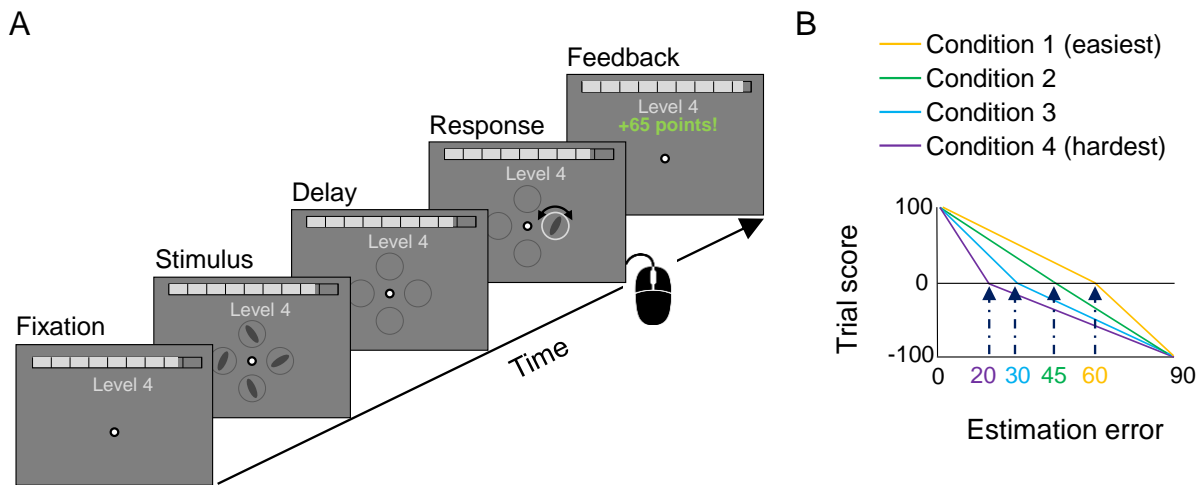


Figure 1 | Design of Experiment 1. Schematic illustration of a single experimental trial. (B) Scoring functions used in the four gamified conditions.

101
102

¹ The room contained two ceiling-mounted fluorescent lamps. The one farthest away from the experimental setup was turned on and the other on was turned off. However, a number of subjects reported that the lamp closest to the setup had spontaneously turned on during the experiment, possibly due to a technical error in the lamp's motion detector. Moreover, one subject was accidentally tested in a completely darkened room.

103 2.4. Task

104 Subjects performed a delayed estimation task (Wilken & Ma, 2004) with stimulus orientation
105 as the relevant feature. Each trial started with a central fixation mark which the subject had to
106 fixate at for at least 600 ms. If a subject failed to fixate within 1 second, a message would appear
107 on the screen telling the subject to “Please fixate at the central dot”. After successful fixation, the
108 subject was presented with N oriented ellipses (200 ms). After a memory delay (1.5 s), the location
109 of one ellipse was highlighted with a dashed circle. The task of the subject was to reproduce the
110 orientation of the earlier displayed ellipse at that location. In order to respond, the subject would
111 first click anywhere on the dashed circle to indicate their initial orientation estimate. Next, an
112 ellipse would appear whose orientation lined up with where the subject had clicked the circle. The
113 subject could then adjust the orientation by moving the mouse and submit the response by clicking
114 the left mouse button. The inter-trial time was 700 ms.

115 *Gamified conditions.* To gamify the task, we extended it with two of the most commonly
116 used concepts in gamification (Hamari et al., 2014): points and levels. We equated the level of the
117 “game” with the set size, i.e., the number of elements in the stimulus display. Subjects started at
118 level 1 and would level up to the next set size each time they had accumulated 1,500 points². On
119 each trial, they earned or lost points based on the accuracy of their response. The scoring function
120 was a two-part linear function in which the minimum error (0°) mapped to 100 points and the
121 maximum error (90°) to -100 points. The error magnitude that mapped to 0 points, referred to as
122 E_0 , differed between the four gamified conditions ($E_0=20, 30, 45, 60$) and determined the difficulty
123 of the task: the higher E_0 , the easier it was to score points (Figure 1B). We expected that task
124 difficulty might influence intrinsic motivation levels (e.g., lower motivation when the task was
125 considered too easy or too hard). Throughout the experiment, a progress bar was visible at the top
126 of the screen to indicate how far or close the subject was to reaching the next level (Figure 1A).
127 After each trial, the subject received feedback about the number of gained or lost points and saw
128 the progress bar grow or shrink accordingly. Visualizing subjects’ competence may be expected
129 to further increase their intrinsic motivation for performing the task (Ryan & Deci, 2000). The

² More precisely, the level was computed at the beginning of each trial as $\left\lfloor \frac{\text{score}}{1500} \right\rfloor + 1$, where the
[] operation rounds a number down to the closest integer

130 gamified conditions consisted of two rounds, each starting at level 1 and lasting 30 minutes.
131 Having two rounds allowed us to examine learning effects and to dissociate these effects from
132 potential other effects. After each round, subjects were told on the screen that they could either
133 “Perform an additional 20 trials to improve your score” or “Skip the extra trials”. The choice that
134 subjects made here was used as “free choice” measure of motivation (Deci, 1971, 1972).

135 *Control condition.* In the control condition, no information about leveling or scores was
136 present on the screen or in the instructions. Instead of leveling up based on performance, subjects
137 progressed from one set size to another after a fixed amount of time. Specifically, they spent 178,
138 204, 260, 354, 370, and 436 seconds on set sizes 1, 2, ..., and 6, respectively. These durations were
139 chosen to match the mean duration that subjects in the $E_0=45^\circ$ gamified group spent on the same
140 set sizes. Hence, the trial progression in the control condition roughly matched the trial progression
141 in one of the gamified conditions, but without the presence of any gamified elements. Just as in
142 the gamified conditions, subjects selected at the end of each run whether they wanted to perform
143 20 extra trials, but without a mentioning of score improvement.

144

145 **2.5. Procedure**

146 Subjects completed the entire experiment in a single session of approximately 90 minutes.
147 After receiving general information about the experiment and signing an informed consent form,
148 they received specific instructions about the task. Subjects in the gamified conditions were told
149 that they would play two rounds of a memory game with the goal to proceed to a level as high as
150 possible. Subjects in the control group were only told that they were going to perform two runs of
151 a memory task with set sizes 1 to 6. Thereafter, the subject performed five practice trials. Control
152 subjects performed the practice trials at set size 1, while subjects in the gamified group performed
153 the trials starting at level = 2.94, to demonstrate the concept of earning points and leveling up.
154 After completing the practice trials, the experiment leader would leave the room and the subject
155 would start the first round of the experiment. After the first round, there was a short break and the
156 subject would start the next round without intervention of the experiment leader. After finishing
157 the second round, the experiment leader would return to the room and conduct a questionnaire,
158 which we describe next.

159

160

161 **2.6. Questionnaire**

162 We designed a custom questionnaire to obtain insight into aspects related to a subject's
163 motivation (a copy of it can be found at <https://osf.io/gb2kd/>). The first part consisted of items
164 similar to the ones found in the Intrinsic Motivation Inventory (IMI; McAuley, Duncan, &
165 Tammen, 1989; Ryan, 1982), such as “I found it interesting” and “It was important for me to
166 perform well”. These items measured motivation in four categories: Interest (items 1, 3, 6, and 8),
167 Perceived Competence (items 5 and 9), Pressure/Tension (items 4 and 10), and Effort/Importance
168 (items 2, 7, and 11). All items were rated on a 1 to 7 integer scale. The second part of the
169 questionnaire consisted of items probing the subject's mood (“bored”, “frustrated”, etc.) in relation
170 to different set sizes. On hindsight we found no use for the data from the second part and did not
171 include them in the analyses.

172

173 **2.7. Analysis methods**

174 We analyzed the data using Bayesian statistics (Etz & Vandekerckhove, 2018;
175 Wagenmakers, Love, & Marsman, 2018; Wagenmakers, Marsman, et al., 2018). All tests were
176 performed using the JASP software package (JASP Team, 2019) with default prior settings. The
177 subscripts of the Bayes Factors that we report indicate which test was used: BF_{10} indicates the
178 probability of the data under the alternative hypothesis relative to the probability of the data under
179 the null hypothesis; BF_{+0} indicates the probability of the data under the hypothesis that group 1
180 has a larger mean than group 2, relative to the probability of the data under the null hypothesis;
181 BF_{incl} indicates the probability of the data under models that includes a main effect of the specified
182 factor relative to the probability of the data under models that do not include this main effect. We
183 use the scale provided in Table 1 of Wagenmakers, Love et al. (2018) for interpretation of the
184 strength of evidence (“weak”, “moderate”, etc).

185

186 **2.8. Results**

187 *2.8.1. Analysis of motivation scores.* We first assess whether gamification affected self-
188 reported scores in the motivation categories Interest, Perceived Competence, Pressure/Tension,
189 and Effort/Importance. For each subject, we compute a single score for each category by averaging
190 across all items within that category. We find that on average, subjects in the gamified conditions
191 reported higher scores in all categories than control subjects (Figure 2). Bayesian t-tests reveal

292 extremely strong evidence for a difference in the category of Interest ($BF_{+0}=172$), strong evidence
 293 in the category of Perceived Competence ($BF_{+0}=88.8$), and moderate evidence in the category of
 294 Effort/Importance ($BF_{+0}=3.19$). In the category of Pressure/Tension, there was weak evidence in
 295 favor of the null hypothesis ($BF_{10}=0.32$). Based on Bayesian ANOVAs with task difficulty as a
 296 fixed factor and subject as a random factor, we find for none of the motivation categories evidence
 297 that task difficulty affects the self-reported motivation scores (Interest: $BF_{10}=0.13$; Perceived
 298 Competence: $BF_{10}=0.28$; Pressure/Tension: $BF_{10}=0.64$; Effort/Importance: $BF_{10}=0.18$). In
 299 summary, these data suggest that subjects in the gamified conditions found the task more
 300 interesting, felt more competent, and possibly put more effort into it than control subjects.

301

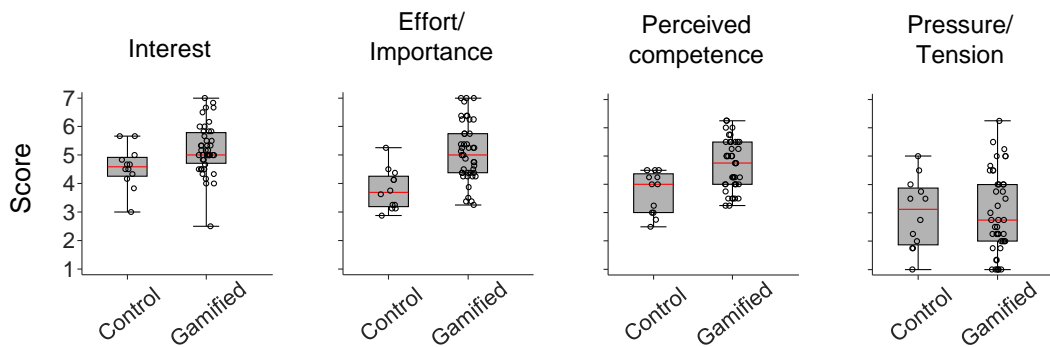


Figure 2 | Motivation scores in Experiment 1. Average self-reported scores on the questionnaire, split by motivation category. The boxes indicate the 25% and 75% quantiles, the red line the median, and the whiskers the most extreme values. The circles indicate scores of individual subjects.

302

303

304 *2.8.2. Effect of task difficulty on VWM performance.* Before we compare performance in the
 305 gamified conditions with performance in the control condition, we examine whether there is a
 306 difference in average performance between the four gamified conditions. To this end, we perform
 307 a Bayesian repeated-measures ANOVA with condition number as a between-subject factor, set
 308 size as a repeated measure, and the circular variance of the estimation error as the dependent
 309 variable. To avoid the results being affected by “survivor bias”, we restrict all analyses of
 310 performance to set sizes for which we have at least 15 measurements for each subject, i.e., set sizes
 311 1 to 4 (at higher set sizes, we lack data for at least one subject in the more difficult conditions).
 312 The result of this test provides moderate evidence for the null hypothesis that there is no difference
 313 in average performance between the gamified conditions ($BF_{incl}=0.22$). Since both the motivation

214 scores and VWM performance seem unaffected by task difficulty, we treat the four gamified
 215 conditions as a single group in the remaining analyses.
 216

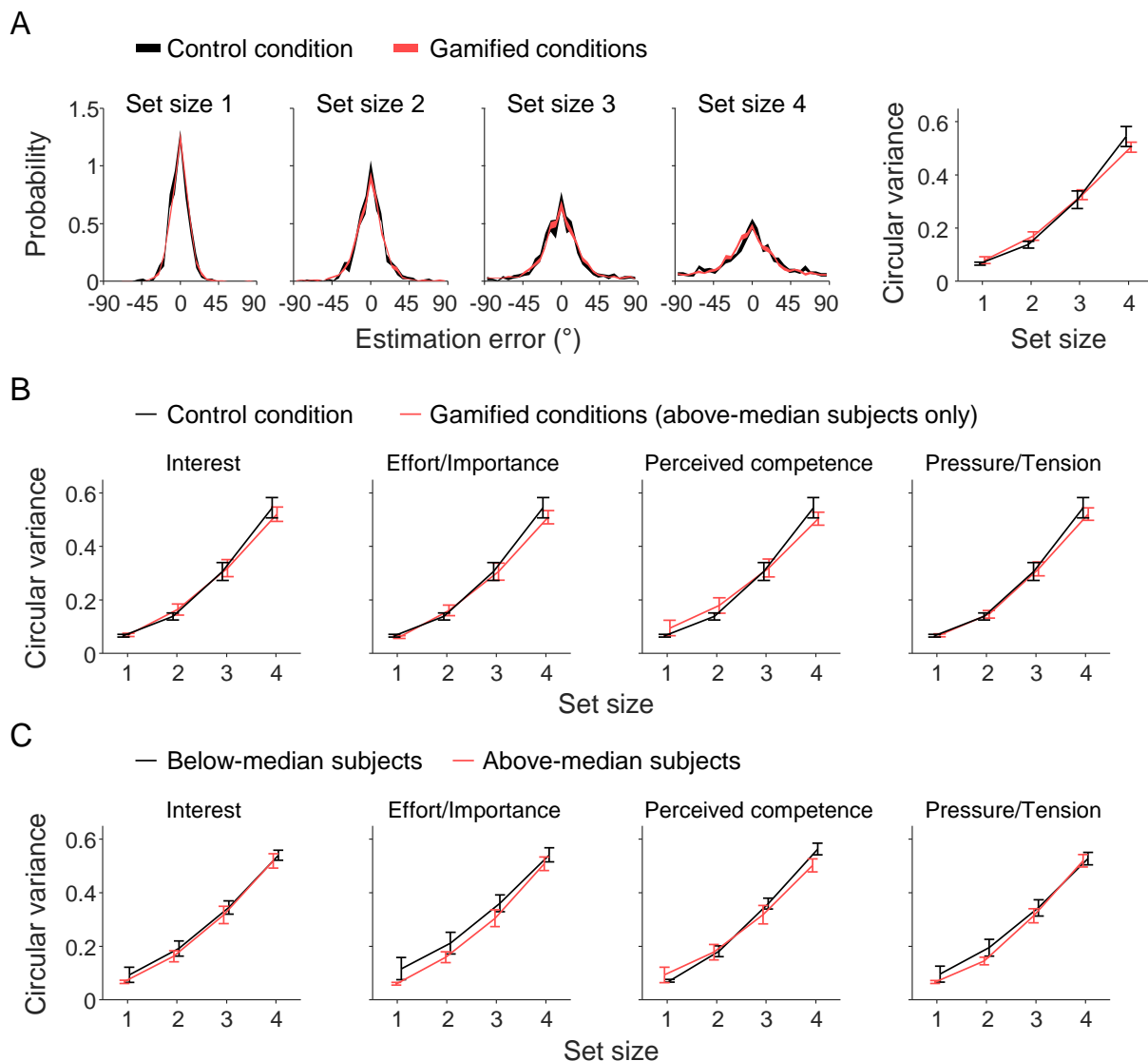


Figure 3 | VWM performance in Experiment 1. (A) Estimation error distributions averaged across subjects in the control condition (black) and subjects in the gamified conditions (red). The width of the curve indicates one standard error. The graph on the right summarizes the histograms by their circular variance. (B) Circular variance compared between subjects in the control condition and subjects with above-median motivation in the gamified conditions. The median split was performed separately for each motivation category. (C) Circular variance compared between below-median and above-median subjects in the gamified conditions.

217
 218

219 2.8.3. *Effect of self-reported motivation on VWM performance.* We next assess whether the
220 motivation differences between the gamified conditions and the control condition are accompanied
221 by differences in VWM performance. The error histograms suggest that this is not the case,
222 because they look virtually identical between subjects in the gamified conditions and subjects in
223 the control condition (Figure 3A). Indeed, a Bayesian repeated-measures ANOVA with set size as
224 a repeated measure, gamification as a binary between-subjects factor, and the circular variance of
225 the error as the dependent measure reveals moderate evidence in favor of the null hypothesis of
226 there being no effect of gamification on VWM performance ($BF_{incl}=0.29$).

227 One possible weakness of the previous analysis is that many subjects in the gamified group
228 had motivation scores similar to the control subjects (Figure 2), which suggests that gamification
229 increased motivation for only part of the subjects. Therefore, a more powerful analysis may be
230 achieved by comparing control subjects with only the highly motivated subjects in the gamified
231 conditions. To this end, we divide subjects in the gamified conditions into below and above-
232 median groups based on the motivation scores. We find no evidence that subjects with an above-
233 median Interest score performed on average differently from the control subjects ($BF_{incl}=0.30$).
234 We neither find evidence for an effect when dividing subjects based on a median split in any of
235 the other three motivation categories (Effort/Importance: $BF_{incl}=0.28$; Pressure/Tension:
236 $BF_{incl}=0.27$; Perceived Competence: $BF_{incl}=0.31$). Finally, we compare the above-median subjects
237 in the gamified conditions with below-median subjects (Figure 3C). Again, regardless of which
238 motivation category we use to make the median split, we find no evidence for a difference in VWM
239 performance between the two groups (Interest: $BF_{incl}=0.26$; Effort/Importance: $BF_{incl}=0.37$;
240 Pressure/Tension: $BF_{incl}=0.26$; Perceived Competence: $BF_{incl}=1.06$).

241 2.8.4. *Effect of free-choice behavior on VWM performance.* We next analyze whether
242 subjects who voluntarily chose to perform additional trials at the end of a round performed better
243 than subjects who chose not to do so. A total of 41 subjects chose to decline the option in both
244 rounds, 11 subjects performed one extra block of trials, and 10 subjects performed both blocks.
245 We perform a Bayesian repeated-measures ANOVA with set size as a within-subject factor, the
246 number of extra blocks that the subject performed as a between-subject factor, and circular
247 variance as the dependent variable. The result reveals moderate evidence against an effect of the
248 number of extra block ($BF_{incl}=0.12$). Hence, even if subjects who voluntarily chose to perform

249 additional trials were more motivated, this was not accompanied by an increase in performance on
250 the VWM task.

251 *2.8.5. Effect of level progress on VWM performance.* So far, we have examined whether
252 there are between-subject differences in performance based on several motivation measures. Next,
253 we assess whether there are any within-subject effects, based on how far a subject had progressed
254 in a level. It could be, for example, that subjects try extra hard when they are close to leveling up
255 or down. For this analysis, we divide the data of subjects in the gamified conditions into 5 bins,
256 with the first bin containing all trials during which the level progress bar was 0-20% filled, the
257 second bin containing all trials during which the bar was filled 20-40%, etc. As before, we only
258 include data from set sizes 1 to 4. A Bayesian ANOVA with bin as a within-subject factor reveals
259 very strong evidence for the null hypothesis of there not being an effect ($BF_{incl}=0.030$).

260 *2.8.6. Effect of round on VWM performance and motivation.* Finally, we examine whether
261 there are any differences in motivation or performance between the two experiment rounds. In the
262 control group, scores on Interest, Effort/Importance, and Perceived Competence dropped by
263 1.31 ± 0.23 , 1.31 ± 0.20 , 1.00 ± 0.26 points, respectively. The score on Pressure/Tension was similar
264 in both rounds (Figure 4A, left). Consistent with the visual impression, Bayesian paired t-tests
265 provide evidence for a drop in the first three categories ($BF_{+0}=432$, $1.08\cdot 10^3$, and 33.7,
266 respectively), but not in the last one ($BF_{+0}=0.23$). To test whether the differences in motivation
267 scores is accompanied by a difference in VWM performance, we perform a two-way Bayesian
268 ANOVA with set size and round number as independent variables and circular variance as the
269 dependent variable. We find no evidence for a difference of round number on VWM performance
270 ($BF_{incl}=0.83$; Fig 4B, left). Indeed, averaged across all control subjects and set sizes, the relative
271 difference in circular variance between the two rounds is nearly zero (they performed $1.1\%\pm 7.2\%$
272 better in the second round).

273 In the group of subjects who performed the gamified version of the experiment, there was
274 no evidence for a drop in Perceived Competence ($BF_{+0}=0.14$) or Pressure/Tension ($BF_{+0}=0.12$)
275 between the first and second round. However, consistent with the control group, there was
276 evidence for a drop in both Interest and Effort/Importance ($BF_{+0}=78.6$ and 48.1, respectively),
277 albeit with much smaller magnitudes (Interest: 0.55 ± 0.15 ; Effort/Importance: 0.44 ± 0.13 ; Figure
278 4B). Interestingly, in contrast to the control group, we find strong evidence for an effect of round
279 on VWM performance in the gamified group ($BF_{incl}=2.35\times 10^5$). Further analysis reveals that the

280 circular variance of the error was $8.5 \pm 4.3\%$ lower in the second round compared to the first round,
 281 indicating an improvement in performance (Figure 4B, right). Since this effect is opposite in
 282 direction from what one would expect as a result of the drops observed in Interest and
 283 Effort/Importance, we believe that it is best interpreted as a learning effect.

284 In summary, the results of the comparison between rounds suggest that gamification helps
 285 to keep subjects more interested and engaged in the task over a longer period of time. Moreover,
 286 we find indications of a learning effect in the gamified conditions, but not in the control condition.
 287 One potential explanation is that the sustained engagement of subjects in the gamified tasks was
 288 beneficial for learning. However, the lack of a learning effect in the control condition may just as
 289 well have been due to lower statistical power (only 12 subjects compared to 50 in the gamified
 290 conditions). Most importantly, consistent with the previous analyses, we find no evidence that
 291 higher motivation is accompanied by better VWM performance.

292

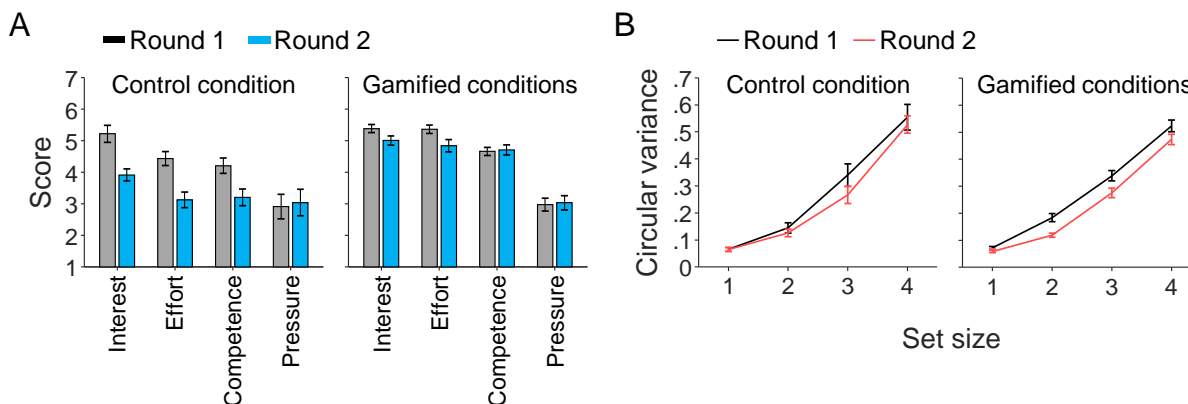


Figure 4 | Motivation and performance differences between rounds in Experiment 1. (A) Difference in self-reported motivation scores between the two rounds. (B) Difference in circular variance of the estimation error distribution between the two rounds.

293

294

295 2.9. Discussion

296 We gamified a standard VWM task and measured how this affected people's motivation
 297 and performance. While the questionnaire data suggest that subjects in gamified conditions had
 298 higher motivation than subjects in the control condition, we did not find any evidence for a
 299 difference in memory performance. This null effect on performance is consistent with earlier
 300 indications that VWM performance is not sensitive to monetary incentives (van den Berg, Zou, &
 301 Ma, 2019; Zhang & Luck, 2011).

302 The perhaps most straightforward explanation for the absence of an effect is that the total
303 amount of invested VWM resource may be fixed, meaning that no matter how hard subjects try,
304 they cannot improve their performance. This would be somewhat surprising, however, because
305 such inflexibility stands in stark contrast to the flexibility that has been found in how subjects
306 distribute VWM resources across a given set of items: when certain items have higher associated
307 reward (Morey, Cowan, Morey, & Rouder, 2011) or are more likely to be probed than others (Bays,
308 2014; Emrich et al., 2017; Yoo, Klyszejko, Curtis, & Ma, 2018), subjects assign more resources
309 to that item compared to the other ones. Moreover, it has been found that subjects can increase
310 performance on cued items without a cost for uncued items (Myers, Chekroud, Stokes, & Nobre,
311 2018) and that specific forms of feedback can improve overall performance on VWM tasks (Adam
312 & Vogel, 2016). Both those findings suggest that it is possible to induce a net increase in utilized
313 VWM resources through experimental manipulations. Also, investing a fixed amount of total
314 resource regardless of the task is suboptimal from a resource-rationality perspective (Van den Berg
315 & Ma, 2018).

316 An alternative explanation of the null effect is that our experimental design might not have
317 been suitable for inducing or detecting motivation-related flexibility in VWM resource investment.
318 One potential problem is that we used very short stimulus times (200 milliseconds), which may
319 have led to incomplete encoding of the items (Bays et al., 2009). As a result, the maximum
320 precision in VWM was possibly limited by the quality of the input rather than by the amount of
321 available VWM resources. Moreover, it is possible that motivation manipulations are only
322 effective when they are administered on a trial-by-trial basis, as suggested by an earlier study on
323 the relation between task preparation and reward (Shen & Chun, 2011). To address these potential
324 weaknesses in the experimental design, we perform a second experiment with a longer stimulus
325 presentation time and a within-subject manipulation of motivation.

326

327 **3. EXPERIMENT 2**

328

329 **3.1. Participants**

330 A total of 12 participants with self-reported normal or corrected-to-normal vision were
331 recruited using posters at various campuses of Uppsala University (8 females; age mean \pm s.e.m. =
332 27.6 \pm 1.39). The study was approved by the Regional Ethical Review Board in Uppsala and

333 conducted according to the Declaration of Helsinki Principles. All participants signed informed
334 consent and received a cinema voucher with a value of approximately \$12 for their participation.
335 One of the participants (S6) had earlier participated as a control subject in Experiment 1.

336

337 **3.2. Experiment procedure**

338 The methods of Experiment 2 were identical to those of Experiment 1, except for the
339 following differences. Most importantly, each subject was tested under three different scoring
340 function with varying amounts of maximum reward (Figure 5B). The scoring functions were of
341 the form $\text{score} = A \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$, where ε is the estimation error in degrees, A determines the
342 maximum score (obtained when $\varepsilon=0$), and σ determines how quickly the score declines as a
343 function of the error. The “low reward” function always gave 7 points, regardless of the accuracy
344 of the subject’s response ($A=7, \sigma=\infty$), the “medium reward” function gave a score between 0 and
345 21 ($A=21, \sigma=30$), and the “high reward” function gave a score between 0 and 101 ($A=101, \sigma=20$).
346 Subjects were told prior to the experiment that there were trials with low, medium, and high
347 potential reward. However, to avoid that they would challenge themselves to always score the
348 maximum number of points – regardless of the scoring function – we did not tell them the
349 maximum score associated with each function. Within each block of 15 trials, each scoring
350 function was used five times, presented in a random order. To inform the subjects about the reward
351 level of the upcoming trial, we added a reward cue (1.1 sec) at the start of the trial (Figure 5A).

352 We made several other minor changes to the experiment. First, we increased the stimulus
353 presentation time from 200 to 900 ms, to reduce the risk that memory quality is limited by the
354 quality of the sensory input rather than the availability of VWM resources. Second, to collect more
355 trials per set size, we increased the number of points required for levelling up from 1,500 to 1,750.
356 Third, we made a minor improvement in how we controlled for fixation errors. In addition to
357 ensuring that subjects were fixating at the start of the trial, we now also forced them to keep
358 fixating during the stimulus presentation and memory delay period. If they broke fixation³ during
359 this period, the trial would be terminated with a message “Invalid trial. Please fixate at the center
360 when the stimulus is shown” or “Invalid trial. Please fixate at the center until the response stage”.

³ We considered fixation to be broken once the eye tracker returned 5 measurements in which the gaze location was farther than 150 pixels away from the center of the fixation dot.

361 The time spent on invalid trials was added to the round time, such that each subject spent 30
 362 minutes on valid trials, regardless of the number of invalid trials. Fourth, we added a 30-second
 363 forced break between set sizes; after those 30 seconds, subjects could resume the experiment by a
 364 keypress whenever they felt ready. Finally, we increased the number of practice trials from 5 to 7
 365 to demonstrate what would happen when breaking of fixation.
 366

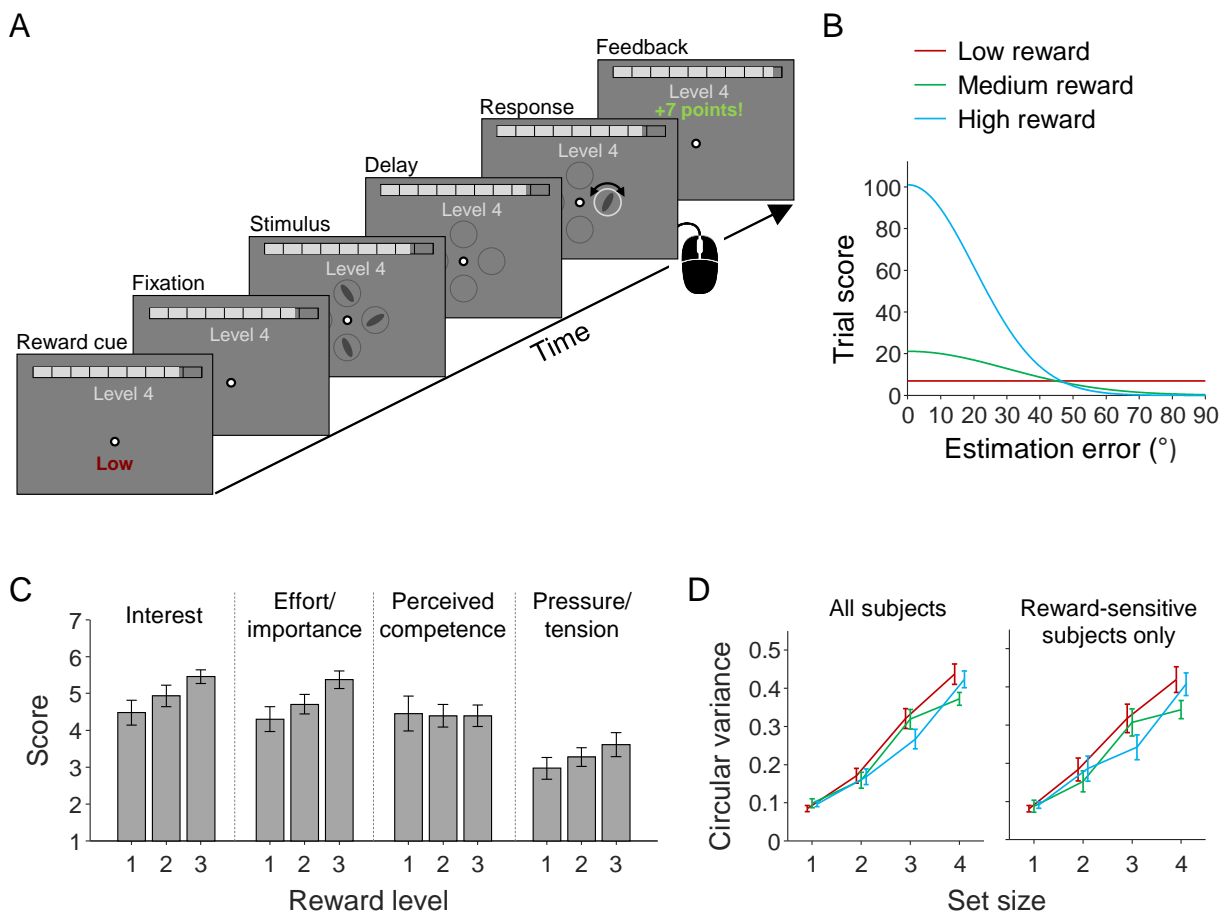


Figure 5 | Design and results of Experiment 2. (A) Schematic illustration of a single experimental trial. (B) Scoring functions used in the three trial types. (C) Motivation scores split by category and reward level. (D) Circular variance of the estimation error distribution split by reward level (see panel B for legend), plotted separately for all subjects (left) and for the group of subjects who reported higher motivation for high-reward trials (right).

367
 368
 369 **3.3. Questionnaire**
 370 After finishing the experiment, subjects filled out a questionnaire in a web browser. The
 371 questions were the same as in the first part of the questionnaire used in Experiment 1, except that

372 a few items were added in each of the four categories (again based on IMI). For all the items we
373 asked separate ratings for low, medium, and high reward trials (a copy of the questionnaire can be
374 found at <https://osf.io/gb2kd/>). Items 1, 3, 6, 8, 13, and 15 were used to compute Interest scores,
375 items 2, 7, 12, and 17 to compute Effort/Importance scores, items 5, 9, 14, and 16 to compute
376 Perceived Competence scores, and items 4, 10, 11 to compute Pressure/Tension scores. The second
377 part of the questionnaire of Experiment 1 was not included.

378

379 **3.4. Results**

380 A visual inspection of the questionnaire data (Figure 5C) suggests a positive relationship
381 between the reward level (low, medium, high) of a trial and the average motivation score in the
382 categories Interest, Effort/Importance, and Pressure/Tension. To test this statistically, we perform
383 a Bayesian linear regression with Reward Level (coded as 1, 2, or 3) as a covariate and motivation
384 score as the dependent variable. We find moderate evidence for a positive relation in categories
385 Interest ($BF_{10}=5.03$; 95% credible interval on the coefficient: [0.00, 0.46]) and Effort/Importance
386 ($BF_{10}=3.36$, 95% credible interval on the coefficient: [0.00, 0.70]), but not in categories Perceived
387 Competence ($BF_{10}=0.32$) and Pressure/Tension ($BF_{10}=0.82$). Hence, on average, subjects
388 apparently found the task more interesting at high-reward trials compared to low-reward trials and
389 were willing to put more effort into those trials.

390 We next examine whether the effect of reward level on motivation was accompanied by an
391 effect on VWM performance (Figure 5D, left). We perform a Bayesian ANOVA with circular
392 variance of the error distribution as the dependent variable, set size and reward level as fixed
393 factors, and subject number as a random factor. Just as in the analyses of Experiment 1, we only
394 include set sizes 1 to 4. The results provide strong evidence for the null hypothesis that there is no
395 effect of reward level on the circular variance ($BF_{incl}=0.091$).

396 Closer inspection of the questionnaire data reveals that the increase of Interest and
397 Effort/Importance with reward level exists for only 7 of the 12 subjects. The remaining 5 subjects
398 either had a decreasing or non-monotonic pattern in one of the two motivation categories. This
399 indicates that only 7 subjects may have been sensitive to the reward manipulation. To verify that
400 the null result on VWM performance was not due to inclusion of the other 5 subjects in our
401 analysis, we rerun the ANOVA on only the 7 subjects with consistent motivation score patterns

402 (Figure 5D, right). We again find evidence for the null hypothesis that there is no effect of reward
403 level on VWM performance ($BF_{\text{incl}}=0.17$).

404 Another reason for the null result at the group level could be that there is heterogeneity in
405 the effect direction across subjects, such that the effects cancel each other out (e.g., some subjects
406 may have performed better with higher reward, while others might have performed worse). To
407 examine this, we perform a model comparison at the level of individual subjects. The first model
408 is the standard version of the variable-precision model (van den Berg, Awh, & Ma, 2014; van den
409 Berg et al., 2012), which has three free parameters: \bar{J}_1 (encoding precision at set size 1), τ
410 (controlling the amount of variability in resource allocation), and α (controlling how encoding
411 precision changes with set size). The relation between set size, N , and encoding precision, \bar{J} , is
412 defined as $\bar{J}(N) = \bar{J}_1 N^\alpha$. We use 5-fold cross validation to compare the fit of this model with a
413 variant in which encoding precision depends on both set size and reward level R (coded as 1, 2,
414 and 3): $\bar{J}(N) = \bar{J}_1 N^\alpha R^\beta$, where β is an additional free parameter. When $\beta=0$, the two models are
415 identical. For subjects with a strong relation between reward level and VWM performance, $\beta=0$
416 will not provide a good fit. For those subjects we should find a strong advantage of the second
417 model over the first one. By contrast, we find that for every subject the cross-validated log
418 likelihood difference between the two models is close to zero (range: 1.05 in favor of the first
419 model to 1.99 points in favor of the second model). In fact, for 10 of the subjects the first model is
420 favored. These results corroborate the earlier evidence suggesting that VWM performance was
421 unaffected by reward level.

422

423 **3.5. Discussion**

424 The purpose of Experiment 2 was to address two potential weaknesses in Experiment 1 by
425 increasing the stimulus presentation time and manipulating motivation on a trial-by-trial basis. The
426 post-questionnaire data of Experiment 2 showed a positive relation between Interest and
427 Effort/Important scores on the one hand and reward level on the other hand, indicating that our
428 within-subject design had the desired effect on motivation. However, the increase in motivation
429 on high-reward trials was not accompanied by an increase in VWM performance. Hence, the
430 results of Experiment 2 are consistent with those of Experiment 1: gamification had a positive
431 effect on motivation, but left performance unaffected.

432 4. GENERAL DISCUSSION

433 While it is known that gamification can increase subjects' intrinsic motivation, enjoyment,
434 engagement, and perceived competence in a task (Hamari et al., 2014), little is known about effects
435 of gamification on cognitive performance. Here, we performed two experiments to investigate
436 whether gamification of a standard VWM task improves subjects' performance on that task.
437 Consistent with previous literature, we found positive effects of gamification on subjects' self-
438 reported interest in the task and the amount of effort they reportedly put into it. However, in neither
439 experiment did we find any differences in VWM performance. We thus conclude that gamification
440 can make people more motivated to perform VWM tasks, but it does not necessarily make them
441 better at it.

442 Our finding that there was no relation whatsoever between motivation and VWM
443 performance in our experiments is puzzling for several reasons. First, there is convincing evidence
444 that people are able to flexibly assign more VWM resources to important items compared to less
445 important ones, in a way that seemingly optimizes performance (Bays, 2014; Emrich et al., 2017;
446 Morey et al., 2011; Yoo et al., 2018). As we have demonstrated earlier (van den Berg & Ma, 2018),
447 it would be suboptimal if people would always invest the same amount of VWM, independent of
448 task properties. Our finding that they nevertheless seemed to do so in our experiments raises the
449 question why VWM has evolved to be approximately optimal in distributing a given amount of
450 resource across items, but at the same time uses a highly suboptimal policy to determine the total
451 amount of VWM resource to invest in a task. Moreover, it is known that motivation and expected
452 reward affect dopamine concentration (e.g., Di Chiara, 2005; Hamid et al., 2015; Syed et al., 2015)
453 and that dopamine concentration, in turn, is related to VWM performance (e.g., Arnsten, 1998;
454 Ashby & Valentin, 2017; Cools et al., 2008; Okimura et al., 2015; Sawaguchi & Goldman-Rakic,
455 1991; Williams & Goldman-Rakic, 1995). Based on those findings, one would expect – contrary
456 to our findings – that any experimental manipulation that affects participants' motivation, also
457 affects their VWM performance.

458 An alternative explanation for our null findings is that subjects may have been
459 overperforming: regardless of the condition they were in, they may have felt a responsibility to
460 perform as well as they could, for instance as a justification for their payment, out of a desire to
461 deliver high-quality data to the researchers, or out of fear to be confronted with their performance
462 after finishing the experiment. Based on informal feedback, we know that at least one of the

463 subjects who voluntarily performed the additional trials did so because she believed it would help
464 the researchers. While most studies that manipulate motivation do so by incentivizing subjects, if
465 overperformance is a real issue, it might be interesting for future studies to look into ways to
466 *decrease* motivation levels. Moreover, it may be that highly artificial stimuli are not very suitable
467 for manipulating motivation levels, for instance due to a lack of engagement in the task. Therefore,
468 it might also be fruitful if future studies would use more naturalistic stimuli. Another potential
469 limitation of the present study is that we used a questionnaire to measure how motivated subjects
470 were and how much effort they put into the task. Another limitation of the present study is that we
471 only analyzed VWM performance at relatively low set sizes (1 to 4). We excluded higher set sizes
472 in order to avoid “survivor bias” effects in the results. However, the subjects’ answers to the
473 questionnaire items that we used to measure intrinsic motivation were based on their experience
474 of all set sizes they performed. Hence, the motivation data may be unreliable for the set sizes that
475 we used to analyze VWM performance differences. To address this problem, it would have been
476 good to complement the self-report measures with pupillary dilation data, which is known to
477 correlate with attention and effort (Hoeks & Levelt, 1993; van der Wel & van Steenbergen, 2018).
478 In the present study we collected eye movement data, including pupil dilation measurements, but
479 we did so mainly to verify that subjects were fixating and not with the aim to use them in our
480 analyses. As a result, the collected data are unfortunately unsuitable for analysis.

481 As a final remark, we believe that our study may serve as a showcase of open science in a
482 field that is currently plagued by concerns about replicability (Aarts et al., 2015) . Although an
483 increasing number of authors make their data available – partly driven by changes in journal
484 policies – researchers still seem to be wary of publishing null results, especially when this a result
485 contradicts their own theory. Keeping such findings hidden in a file drawer may have short-term
486 benefits for the researchers, but is seen as one of the major factors behind the replicability crisis.

487

488 **ACKNOWLEDGEMENTS**

489 RvdB is supported by Grant 2018-01947 from the Swedish Research Council (Vetenskapsrådet)
490 and Grant RIK17-1157:2 of the Bank of Sweden Tercentenary Foundation (Stiftelsen Riksbankens
491 Jubileumsfond). The funding sources had no role in the study design, the collection, analysis and
492 interpretation of data, the writing of the report, and the decision to submit the article for
493 publication.

494 **DECLARATION OF INTEREST**

495 None.

496

497 **REFERENCES**

498 Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., ... Zuni, K.
499 (2015). Estimating the reproducibility of psychological science. *Science*.

500 <https://doi.org/10.1126/science.aac4716>

501 Adam, K. C. S., & Vogel, E. K. (2016). Reducing failures of working memory with performance
502 feedback. *Psychonomic Bulletin & Review*, 23(5), 1520–1527.

503 <https://doi.org/10.3758/s13423-016-1019-4>

504 Arnsten, A. F. T. (1998). Catecholamine modulation of prefrontal cortical cognitive function.

505 *Trends in Cognitive Sciences*. [https://doi.org/10.1016/S1364-6613\(98\)01240-6](https://doi.org/10.1016/S1364-6613(98)01240-6)

506 Ashby, F. G., & Valentin, V. V. (2017). Multiple systems of perceptual category learning:

507 Theory and cognitive tests. In *Handbook of categorization in cognitive science, 2nd ed.* (pp.

508 157–188). <https://doi.org/10.1016/B978-0-08-101107-2.00007-5>

509 Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the
510 brain. *Journal of Cerebral Blood Flow and Metabolism : Official Journal of the*

511 *International Society of Cerebral Blood Flow and Metabolism*, 21(10), 1133–1145.

512 <https://doi.org/10.1097/00004647-200110000-00001>

513 Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *The*

514 *Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 34(10),

515 3632–3645. <https://doi.org/10.1523/JNEUROSCI.3204-13.2014>

516 Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is
517 set by allocation of a shared resource. *Journal of Vision*, 9(10), 7.1-11.

518 <https://doi.org/10.1167/9.10.7>

519 Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond
520 individual items and toward structured representations. *Journal of Vision*, 11(5), 4.

521 <https://doi.org/10.1167/11.5.4>

522 Cools, R., Gibbs, S. E., Miyakawa, A., Jagust, W., & D'Esposito, M. (2008). Working memory
523 capacity predicts dopamine synthesis capacity in the human striatum. *The Journal of*

524 *Neuroscience : The Official Journal of the Society for Neuroscience*, 28(5), 1208–1212.

525 <https://doi.org/10.1523/JNEUROSCI.4475-07.2008>

526 Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of*
527 *Personality and Social Psychology*. <https://doi.org/10.1037/h0030644>

528 Deci, E. L. (1972). Intrinsic motivation, extrinsic reinforcement, and inequity. *Journal of*
529 *Personality and Social Psychology*. <https://doi.org/10.1037/h0032355>

530 Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to
531 gamefulness: Defining “gamification.” *Proceedings of the 15th International Academic*
532 *MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*.
533 <https://doi.org/10.1145/2181037.2181040>

534 Di Chiara, G. (2005). Chapter VI Dopamine, motivation and reward. *Handbook of Chemical*
535 *Neuroanatomy*. [https://doi.org/10.1016/S0924-8196\(05\)80010-7](https://doi.org/10.1016/S0924-8196(05)80010-7)

536 Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in Education: A
537 Systematic Mapping Study. *Educational Technology & Society*, 18(3), 75–88.
538 <https://doi.org/10.1109/EDUCON.2014.6826129>.

539 Emrich, S. M., Lockhart, H. A., & Al-Aidroos, N. (2017). Attention mediates the flexible
540 allocation of visual working memory resources. *Journal of Experimental Psychology:*
541 *Human Perception and Performance*, 43(7), 1454–1465.
542 <https://doi.org/10.1037/xhp0000398>

543 Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian Inference for Psychology.
544 *Psychonomic Bulletin and Review*. <https://doi.org/10.3758/s13423-017-1262-3>

545 Fuster, J. M., & Alexander, G. E. (1971). Neuron Activity Related to Short-Term Memory.
546 *Science*, 173(3997), 652–654. <https://doi.org/10.1126/science.173.3997.652>

547 Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of
548 empirical studies on gamification. *Proceedings of the Annual Hawaii International*
549 *Conference on System Sciences*. <https://doi.org/10.1109/HICSS.2014.377>

550 Hamari, J., & Lehdonvirta, V. (2010). Game design as marketing: How game mechanics create
551 demand for virtual goods. *International Journal of Business Science and Applied*
552 *Management*.

553 Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C.
554 M., ... Berke, J. D. (2015). Mesolimbic dopamine signals the value of work. *Nature*
555 *Neuroscience*. <https://doi.org/10.1038/nn.4173>

556 Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A
557 longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and
558 academic performance. *Computers and Education*, 80, 152–161.
559 <https://doi.org/10.1016/j.compedu.2014.08.019>

560 Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: a quantitative
561 system analysis. *Behavior Research Methods, Instruments, & Computers*.
562 <https://doi.org/10.3758/BF03204445>

563 JASP Team. (2019). JASP (Version 0.10.2). [Computer Software].

564 Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory
565 information. *Current Opinion in Neurobiology*, Vol. 11, pp. 475–480.
566 [https://doi.org/10.1016/S0959-4388\(00\)00237-3](https://doi.org/10.1016/S0959-4388(00)00237-3)

567 Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and
568 neurobiology to individual differences. *Trends in Cognitive Sciences*, 17(8), 391–400.
569 <https://doi.org/10.1016/j.tics.2013.06.006>

570 Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature*
571 *Neuroscience*, 17(3), 347–356. <https://doi.org/10.1038/nn.3655>

572 McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric Properties of the Intrinsic
573 Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis.
574 *Research Quarterly for Exercise and Sport*, 60(1), 48–58.
575 <https://doi.org/10.1080/02701367.1989.10607413>

576 Morey, C. C., Cowan, N., Morey, R. D., & Rouder, J. N. (2011). Flexible attention allocation to
577 visual and auditory working memory tasks: Manipulating reward induces a trade-off.
578 *Attention, Perception, and Psychophysics*. <https://doi.org/10.3758/s13414-010-0031-4>

579 Myers, N. E., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2018). Benefits of flexible
580 prioritization in working memory can arise without costs. *Journal of Experimental*
581 *Psychology: Human Perception and Performance*. <https://doi.org/10.1037/xhp0000449>

582 Nassar, M. R. M. R., Helmers, J. C. J. C., & Frank, M. . M. J. (2018). Chunking as a rational
583 strategy for lossy data compression in visual working memory. In *Psychological Review*
584 (Vol. 125). <https://doi.org/10.1037/rev0000101>

585 Okimura, T., Tanaka, S., Maeda, T., Kato, M., & Mimura, M. (2015). Simulation of the capacity
586 and precision of working memory in the hypodopaminergic state: Relevance to

587 schizophrenia. *Neuroscience*, 295, 80–89.
588 <https://doi.org/10.1016/J.NEUROSCIENCE.2015.03.039>

589 Ryan, R.M.; Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation,
590 social development, and well-being. *American Psychologist*.

591 Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of
592 cognitive evaluation theory. *Journal of Personality and Social Psychology*.
593 <https://doi.org/10.1037/0022-3514.43.3.450>

594 Sawaguchi, T., & Goldman-Rakic, P. (1991). D1 dopamine receptors in prefrontal cortex:
595 involvement in working memory. *Science*. <https://doi.org/10.1126/science.1825731>

596 Shen, Y. J., & Chun, M. M. (2011). Increases in rewards promote flexible behavior. *Attention,*
597 *Perception, and Psychophysics*. <https://doi.org/10.3758/s13414-010-0065-7>

598 Sterling, P., & Laughlin, S. (2015). *Principles of neural design*. MIT Press.

599 Subhash, S., & Cudney, E. A. (2018). Gamified learning in higher education: A systematic
600 review of the literature. *Computers in Human Behavior*.
601 <https://doi.org/10.1016/j.chb.2018.05.028>

602 Syed, E. C. J., Grima, L. L., Magill, P. J., Bogacz, R., Brown, P., & Walton, M. E. (2015).
603 Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nature*
604 *Neuroscience*. <https://doi.org/10.1038/nn.4187>

605 van den Berg, R., Awh, E., & Ma, W. J. W. J. (2014). Factorial comparison of working memory
606 models. *Psychological Review*, 121(1), 124–149. <https://doi.org/10.1037/a0035234>

607 van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in visual
608 working memory. *ELife*, 7(e34963). <https://doi.org/10.7554/eLife.34963>

609 van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in
610 encoding precision accounts for visual short-term memory limitations. *Proceedings of the*
611 *National Academy of Sciences*, 109(22), 8780–8785.
612 <https://doi.org/10.1073/pnas.1117465109>

613 van den Berg, R., Zou, Q., & Ma, W. J. (2019). Does monetary reward increase visual working
614 memory performance? *BioRxiv*, 767343. <https://doi.org/10.1101/767343>

615 van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive
616 control tasks: A review. *Psychonomic Bulletin and Review*. [https://doi.org/10.3758/s13423-](https://doi.org/10.3758/s13423-018-1432-y)
617 [018-1432-y](https://doi.org/10.3758/s13423-018-1432-y)

618 Wagenmakers, E. J., Love, J., & Marsman, M. (2018). Bayesian inference for psychology. Part
619 II: Example applications with JASP. *Psychon Bull Rev*, 25(1), 58–76.

620 Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... Morey, R. D.
621 (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical
622 ramifications. *Psychonomic Bulletin and Review*. [https://doi.org/10.3758/s13423-017-1343-](https://doi.org/10.3758/s13423-017-1343-3)
623 3

624 Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of*
625 *Vision*, 4(12), 1120–1135. <https://doi.org/10.1167/4.12.11>

626 Williams, G. V., & Goldman-Rakic, P. S. (1995). Modulation of memory fields by dopamine D1
627 receptors in prefrontal cortex. *Nature*. <https://doi.org/10.1038/376572a0>

628 Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working
629 memory resource. *Scientific Reports*. <https://doi.org/10.1038/s41598-018-34282-1>

630 Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory.
631 *Psychological Science*, 22(11), 1434–1441. <https://doi.org/10.1177/0956797611417006>

632