

1 Division of labour biased towards equality amongst 2 collaborators in dyadic short-term memory task

3

4 Edwin S. Dalmaijer ^{1,2}, Holly Knapton ³, Masud Husain ¹, Kenneth Holmqvist ⁴, & Diederick C.
5 Niehorster ^{3,5}

6

7 *1. Department of Experimental Psychology and Wellcome Centre for Integrative Neuroimaging,
8 University of Oxford, UK*

9 *2. MRC Cognition and Brain Sciences Unit, University of Cambridge, UK*

10 *3. Department of Psychology, Lund University, Sweden*

11 *4. Department of Psychology, Regensburg University, Germany; Universiteit van die Vrystaat,
12 Bloemfontein, South Africa; UPSET, NWU Vaal, South Africa; Faculty of Arts, Masaryk University,
13 Brno, Czech Republic*

14 *5. Humanities Laboratory, Lund University, Sweden*

15

16 Corresponding author

17 Edwin Dalmaijer (edwin.dalmaijer@mrc-cbu.cam.ac.uk), MRC Cognition and Brain Sciences Unit,
18 15 Chaucer Road, CB2 7EF, Cambridge, United Kingdom.

19

20 Manuscript details

21 Abstract word count: **169**

22 Word count (excluding captions):

23 Introduction: 1307

24 Methods: 2824

25 Results: 1220

26 Discussion: 876

27 **Total:** **6227**

28 Tables: 0

29 Figures: 7

30

31 Acknowledgements

32 ESD was funded through Marie Curie grant 606901 (“INDIREA”) from the European Commission.
33 MH is funded by the Wellcome Trust. The Wellcome Centre for Integrative Neuroimaging is
34 supported by core funding from the Wellcome Trust (203139/Z/16/Z). The authors gratefully
35 acknowledge Lund University Humanities Lab. Finally, we would like to thank the asbestos-related
36 sudden closure of Oxford’s Tinbergen Building, which allowed ESD just enough time to visit the
37 unique facilities managed by DCN, where the current study was developed and run.

38

39 **Abstract**

40

41 Human workers often divide labour on the basis of individual qualities. However, contemporary
42 work on decision-making suggests collaborators value opinions equally, even when they are
43 unequally skilled. We investigated whether this equality bias extends to division of labour in a
44 collaborative game that leverages inter-individual short-term memory capacity differences. Pairs
45 were shown eight memorisable items to divide among themselves, only communicating through
46 their computer screens that displayed who claimed which items. After this, participants recalled
47 randomly selected claimed items. To incentive collaboration, pairs were rewarded for their
48 combined recall accuracy. Although we hypothesised they would maximise reward by dividing
49 items according to each individuals' capacities, pairs divided the number of items to-be-
50 remembered equally. Furthermore, individuals' collaboration ratings were unaffected by capacity or
51 performance, but instead negatively correlated with task-irrelevant variability in claimed item
52 locations. Finally, differences in claimed item numbers correlated with inter-individual differences
53 in conscientiousness and social apathy. Our findings suggest humans have an equality bias when
54 dividing labour between collaborators, even if they are of unequal skill.

55

56 **Keywords:** social cognition; collaboration; apathy; motivation; visual short-term memory;
57 behavioural volatility; optimal behaviour

58

59

Introduction

60

61 Human societies are built on collaborative behaviour that varies in scope from organisations
62 consisting of couples to many thousands of people working towards a shared goal. Interestingly,
63 verbal communication is not a pre-requisite for successful collaboration. The human ability to
64 collaborate on the basis of very little information is illustrated by the first study of its kind, in which
65 pairs of collaborating individuals (also known as “dyads”) were asked to jointly search for a
66 particular target (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008). The individuals, each in a
67 separate room, were presented with the same field of distractor stimuli (“Q”) among which a target
68 (“O”) could be embedded. In one condition, the only information shared between the pair was
69 where each individual was currently looking, which was recorded by an eye tracker, and displayed
70 on the other’s display. Collaborative search proved to be faster than the solitary equivalent.

71 One important observation in the aforementioned study was that pairs implicitly divided the
72 search space in two halves, with each taking responsibility for their own half (Brennan et al., 2008).
73 In a condition in which pairs were allowed to speak, pairs verbally decided on a highly similar
74 strategy, although being allowed to communicate in this way did not seem to improve performance
75 (Brennan et al., 2008). This finding has since been replicated (Neider, Chen, Dickinson, Brennan, &
76 Zelinsky, 2010). In addition, the collaborative benefit seems to correlate with the similarity of
77 individuals’ verbal communication style, and their social proximity (Brennan & Enns, 2015).
78 Furthermore, the benefit of knowing about collaborators’ actions extends across perceptual
79 domains, for example when this information is auditory or even vibro-tactile (Wahn et al., 2016).

80 Modelling work suggests that joint visual search with real-time gaze feedback is quicker
81 because of active collaboration (Brennan & Enns, 2015), but not when participants simply follow
82 each other’s gaze around the search array (Messmer, Leggett, Prince, & McCarley, 2017; Yamani,

83 Neider, Kramer, & McCarley, 2017). The latter behaviour occurs more in competitive joint search,
84 whereas collaborative search facilitates participants avoiding each other's gaze and thus more
85 efficiently dividing the search space (Niehorster, Cornelissen, Holmqvist, & Hooge, submitted).
86 When joint visual search is performed without direct feedback, collective search is also faster than
87 independent search, albeit to a lesser extent. Specifically, when feedback is given only after a trial,
88 performance is better for groups of two or three compared to single participants, and is correlated
89 with individual search speed (Wahn, Czeszumski, & König, 2018). This suggests that without real-
90 time feedback, a collective benefit can be achieved through statistical facilitation rather than
91 collaboration. However, real-time information on collaborators improves performance to a greater
92 extent. In the literature, these concepts are referred to as "collective benefit" and "collaborative
93 benefit" (Wahn, Kingstone, & König, 2018).

94 Behavioural adaptation to group performance

95 Although the summarised (simple) cognitive tasks require individuals to coordinate or
96 respond to each other's actions, they do not require individuals to be aware of each other's cognitive
97 skills. By contrast, real-life collaborations often require a division of labour that is in keeping with
98 individual capacity. For example, when lifting wooden planks of increasing length, pairs of
99 participants decide to start jointly carrying planks at a length that was proportional to the pair's
100 mean arm span (Richardson, Marsh, & Baron, 2007; Sebanz, Bekkering, & Knoblich, 2006). This
101 illustrates that pairs were not only aware of each other's capacity to carry planks, but also able to act
102 on this knowledge.

103 Recent work on joint perceptual decision-making has demonstrated that individuals within a
104 pair assume their collaborator is equally capable, and that violations of this assumption are
105 detrimental to joint performance (Mahmoodi et al., 2015). In this study, participants had to first
106 individually decide on whether a target appeared in a first or second interval, and in the case of

107 disagreement one was randomly chosen to be arbitrator and make a final decision. Crucially, each
108 individual provided a confidence rating, which the arbitrator could use to make their decision.
109 Arbitrators were shown to weigh each response equally, even when differences in task performance
110 were introduced between the participants by making one individual's task harder, and when
111 participants received a group reward (Mahmoodi et al., 2015). This elegantly suggests that humans
112 weigh their collaborators' opinions equally rather than optimally.

113 Unfortunately, perhaps in a small-scale application of the Dunning-Kruger effect (Kruger &
114 Dunning, 1999), less skilled individuals were more likely to report higher confidence in wrong
115 decisions (Mahmoodi et al., 2015). This is an inherent problem with self-report, and potentially
116 clouded arbitrators' ability to make a skill-based decision in trials with disagreement. In addition,
117 individual participants could not directly affect the performance of the other participant. Hence, it
118 remains unclear whether humans would act on the perceived cognitive skills of collaborators when
119 they are given a chance to do so.

120 Joint visual short-term memory

121 Here, we present a novel collaborative task that requires a pair of participants to remember a
122 number of visual stimuli. In one variety of visual short-term memory tasks (Zhang & Luck, 2008),
123 participants are presented with a memory array of several items, for example coloured squares or
124 bars with a particular orientation. They are then required to maintain the items in memory for a
125 duration of several seconds. After this delay, one item is probed, and participants are required to
126 reproduce the original colour or orientation.

127 The advantage of using such a task to investigate collaborative behaviour, is that visual
128 short-term memory is a well-studied system (Baddeley & Hitch, 1974; Brady, Konkle, & Alvarez,
129 2011; Luck & Vogel, 2013; Ma, Husain, & Bays, 2014). Although there is debate on the precise
130 architecture, researchers agree that short-term memory is of a limited capacity (Adam, Vogel, &

131 Awh, 2017; Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Baddeley & Hitch, 1974;
132 Bays, 2015; Bays & Husain, 2008; Cowan, 2001; Wilken & Ma, 2004). In addition, visual short-
133 term memory capacity can be measured reliably within individuals, and shows a wide variability
134 between individuals (Luck & Vogel, 2013).

135 In our task, individuals receive objective feedback on their own and their partner's
136 performance. Performance is translated into a group reward in a non-linear way that punishes larger
137 errors. Crucially, participants can claim stimuli for themselves by looking at them, and are later
138 probed only on stimuli that they have claimed. This means that participants can choose to take
139 responsibility for more or fewer stimuli. We hypothesised that participants would learn how well
140 they were doing compared to their partner, and adjust their behaviour accordingly.

141 For example, if a pair of participants is asked to remember a total of eight items, one
142 participant could notice they were doing markedly better than their partner when each claim four
143 items. On the next trial, they could choose to claim one additional item, which would increase their
144 memory load to five items. As a consequence, their partner's memory load would be reduced to
145 three items. Although their own performance would decrease slightly, their partner's performance
146 could increase by a greater amount, and their joint reward would be greater.

147 In addition, we hypothesised that participants would judge the collaboration to be better
148 when their partner was more sensitive to the difference in cognitive ability between them. In other
149 words, we expected pairs that had a larger discrepancy in visual short-term memory capacity to rate
150 their collaboration as better if the higher-capacity individual adjusted their behaviour to help the
151 lower-capacity individual.

152 Finally, we hypothesised that the degree to which people would adjust their behaviour (i.e.
153 take on more or less labour depending on the inter-individual skill difference) would be mediated by
154 basic personality traits like conscientiousness. Furthermore, we had a specific interest in apathy

155 (Ang, Lockwood, Apps, Muhammed, & Husain, 2017), and hypothesised that higher levels of social
156 apathy would correlate with lower levels of behavioural change inspired by inter-collaborator
157 differences in ability.

158

159

160 **Methods**

161

162 **Procedure**

163 After expressing interest, potential participants were approached via email to schedule a
164 testing session. They were also requested to complete an online questionnaire in advance, which
165 included questions on demographic data, Big-Five personality domains (Gosling, Rentfrow, &
166 Swann, 2003), and apathy in the behavioural, social, and emotional domain (Ang et al., 2017).

167 Informed consent to completing the questionnaire was acquired at the start of the online session.

168 Informed consent to participation in the testing session was acquired in written form before the start
169 of the session.

170 Testing sessions took place in Lund University's 'Digital Classroom', which contains 25
171 computers with eye trackers (SMI RED-m at the time of testing) that are connected to the same
172 network, and can share data in real-time. For details, see (Nyström, Niehorster, Cornelissen, &
173 Garde, 2017). Participants were each assigned a cardboard cubicle that contained a desk with a
174 computer and an eye tracker (**Figure 1**). All computers were connected to the same network, and
175 controlled from a master computer.

176



177

Figure 1 – Lund University’s Digital Classroom setup. A total of 25 computers with SMI RED-m eye trackers and a master computer are connected through the same network. Here, we used up to 10 of the computers to run experiments on, and white cardboard cubicles to hide participants from each other’s view.

178

179 During a testing session, participants first completed an individual visual short-term memory
180 assessment (described below). When all participants finished, their eye trackers were calibrated.
181 After calibration, participants performed a collaborative task (described below) with randomly
182 assigned partners. Participants were unaware of who their partner was, as they were not identified
183 within the game. Because participants were inside a cubicle that blocked others from view, and
184 were instructed to remain silent throughout the game, there were no external cues through which
185 partners could be identified.

186 We aimed to run three consecutive collaborative tasks per testing session, with new pairings
187 on every run. We succeeded in this aim in all but one session, in which we did not have time for a
188 third collaborative task run (consenting and calibrating 10 participants took longer than anticipated).
189 Testing sessions did not last longer than 1 hour.

190

191 Participants

192 Participants were recruited through advertisement on informal channels at Lund University,
193 including physical and digital notice boards. In addition, we advertised with the Lund International
194 Society of Psychology (LISP) by offering a free MSc-level lecture on the applications of eye
195 tracking in psychological research (attendants were encouraged but not mandated to participate). In
196 total, 41 individuals expressed interest and completed our pre-participation questionnaires. Out of
197 those, 3 individuals took part in a pilot, and 32 individuals in a testing session. This number
198 includes two authors, who each participated once, in sessions where an odd number of participants
199 showed up. Participants were compensated for their time with gift cards that were valid in a variety
200 of local shops.

201 A total of seven sessions were run. The first (N=3) was a pilot session to test whether the
202 setup worked as expected, and did not yield usable data. The following six sessions (N=9 + author
203 Niehorster, N=4, N=2, N=4, N=4, N=3 + author Dalmaijer) ran smoothly, with the exception of the
204 malfunction of one computer that interrupted two pairs in the first session. In addition, we had to
205 exclude one session where anonymisation was compromised due to only two of the scheduled
206 participants showing up.

207 The testing sessions yielded usable data from 32 unique pairs of individuals. These
208 individuals (N=26, 7 males, 19 females) had a mean age of 25.7 years with a standard deviation of
209 4.1 and a range of 21-38 years, varied in highest completed education from high school to doctorate
210 (2 high school, 14 Bachelor, 7 Master, 3 Doctorate), and had a variety of nationalities (Bulgaria,
211 Germany, Hungary, Iceland, India, Italy, Mexico, Netherlands, Russia, Sweden, Thailand, Turkey,
212 USA).

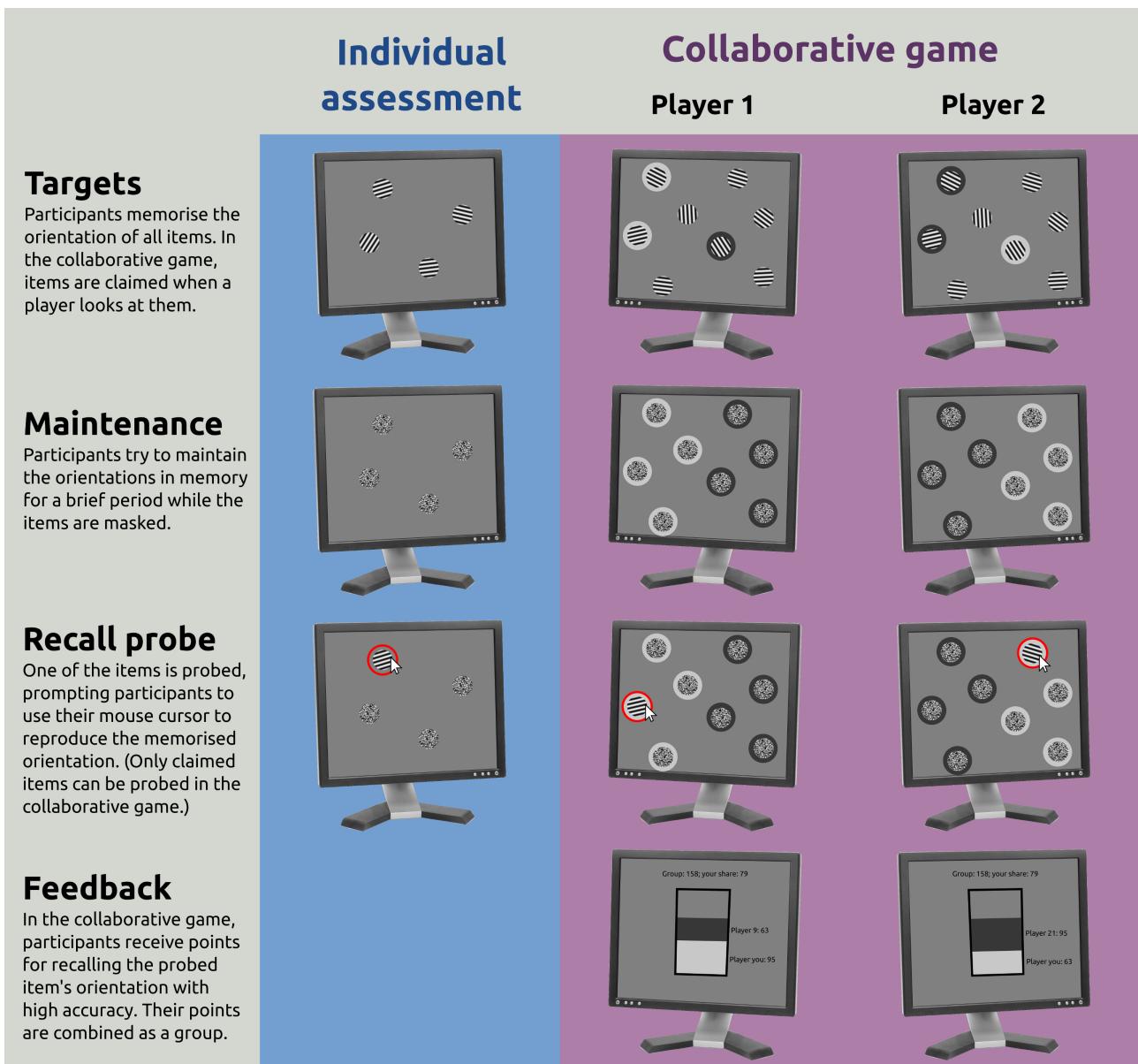
213

214 Individual visual short-term memory capacity assessment

215 The individual short-term memory assessment aimed to test participants' capacity to recall
216 and reproduce visual information (**Figure 2**). It consisted of 80 trials, during each of which 2 or 4
217 items with a unique orientation were presented for 5 seconds. This was followed by a maintenance
218 period of 3 seconds during which the items were masked. The maintenance period ended with the
219 appearance of the mouse cursor, and a circle around one of the (still masked) items. Participants
220 could then use the computer mouse to precisely indicate the orientation of the probed item as they
221 remembered it. On the first click, the mask was replaced by an item that was oriented toward the
222 current mouse position. Participants could hold down the mouse button while moving the cursor to
223 rotate the item to the remembered orientation, and press the space key to confirm.

224 The items were sinusoidal gratings with a diameter of 1.5 degrees of visual angle, and a
225 spatial frequency of five cycles per stimulus. Each item had a unique and randomly generated
226 orientation that differed from the other items by at least ten degrees. Item locations were also
227 randomly generated on each trial, were separated by at least 3.35 degrees of visual angle, and
228 avoided areas near the screen edge (left/right borders were defined as 10% of the screen width, and
229 top/down as 10% of the screen height). Masks were noise patches with the same diameter as the
230 gratings. Each trial was preceded by a centrally presented fixation cross, which was visible for a
231 randomly selected interval between 1 and 1.5 seconds.

232



233

Figure 2 – Experimental procedure in two visual short-term memory tasks. In these tasks, participants see a number of items for a limited amount of time, then hold them in memory for three seconds, after which one of the items is probed. The orientation of the probed item can be adjusted by participants, who are asked to reproduce the memorised orientation. The individual assessment was designed to estimate individual participants' visual short-term memory capacity. The collaborative game was designed to allow participants to divide memory items between them in any way they preferred. Participants received direct visual feedback on what items they and their collaborator claimed.

234

235

We computed errors as the circular distance between a participant's responses and the

236

original orientation of items (i.e. the difference in angle between a probed item and the associated

237 response). For each participant, we established an error distribution for both the two-item and the
 238 four-item condition (40 trials each). These error distributions were fitted (using a maximum-
 239 likelihood procedure) to a mixture model (**Equation 1**) that described the probability density
 240 function of responses, and estimated the proportion of trials in which an item was forgotten β
 241 (guesses), the proportion of trials in which the response to one item was made with another item in
 242 mind γ (“swap errors”), and the precision of recall σ that was defined as the converted spread
 243 parameter (analogous to a normal distributions standard deviation) of a Von Mises distribution
 244 (Bays, Catalao, & Husain, 2009).

245

246
$$p(\hat{\theta}) = (1 - \gamma - \beta)\phi_{\sigma}(\hat{\theta} - \theta) + \gamma \frac{1}{2\pi} + \beta \phi_{\sigma}(\hat{\theta} - \theta^*) \quad (1)$$

247 *Where:*

248 - $\hat{\theta}$ are a participant’s responses on each trial in either the 2-item or 4-item condition

249 - θ are the orientations of the probed item on each trial

250 - θ^* are the orientations of the non-probed items on each trial

251

252 After establishing the proportion of trials in which participants did not guess, and the
 253 precision of their recall in those trials, we could estimate their visual short-term memory capacity κ
 254 (**Equation 2**). Specifically, we computed the Kullback-Leibler divergence (Kullback & Leibler,
 255 1951) between the probability density function that fitted their non-guessed responses best
 256 (**Equation 3**), and the probability density function that describes pure guessing (**Equation 4**). This
 257 metric quantifies the information gain when a participant remembers an item instead of randomly
 258 guessing its orientation. An approximation of visual short-term memory capacity κ (in *bits*) can be

259 computed by multiplying the Kullback-Leibler divergence by the amount of to-be-remembered
 260 items n . Note that here, swap errors are counted towards the computation of memory capacity κ .
 261 Although swap errors originate from information that was bound to the wrong item, this
 262 information was stored in visual short-term memory, and should thus count towards its maximum
 263 capacity.

264

265
$$\kappa = n \int_{-\pi}^{\pi} p(x) \log_2 \frac{p(x)}{q(x)} dx \quad (2)$$

266 Where:

267
$$p(x) = (1 - \gamma) \phi_\sigma(x) + \gamma \frac{1}{2\pi} \quad (3)$$

268

269
$$q(x) = \frac{1}{2\pi} \quad (4)$$

270

271 The final estimate for each participant's visual short-term memory capacity was obtained by
 272 averaging the estimates of κ obtained from the 2-item and 4-item conditions. The associated unit is
 273 bits of information.

274

275 Joint visual short-term memory task

276 We designed a joint short-term memory task that was similar in concept to the individual
 277 task described above (**Figure 2**). In each trial, a pair had to memorise a total of eight items (the
 278 same sinusoidal gratings as in the individual task). After a maintenance period of 3 seconds, each

279 individual was probed to recall one of the items. Responses could again be given by using the
280 mouse to precisely indicate the memorised orientation, as in the individual task. Individual
281 participants within each pair were rewarded with points for their performance (see below for the
282 exact reward scheme). They were shown both their own reward, and the reward that their partner
283 obtained. These rewards were then added together, and equally divided between the participants.
284 Hence, participants were not only incentivised to do well, but also to help their collaborator do well.

285 Crucially, the individuals in each pair were allowed to claim their own share of the items.
286 They could claim an item by looking at an item for 150 ms or longer. After a participant claimed an
287 item, a light shading appeared behind it on their own screen, and a dark shading appeared behind
288 the same item on the other participant's screen. Participants could not claim items that were already
289 claimed. Only claimed items would be probed for a response. If a participant did not claim any
290 items, no item would be probed, and they would be assigned a reward of 0 points. This occurred in
291 only 1 out of all 800 trials. (In this scenario, the other participant would have to claim and
292 remember all eight items.)

293 How many items were claimed by each individual within a pair was determined by the pair.
294 They received an instruction that explicitly informed them that they were allowed to divide the
295 items between themselves and their collaborator in any way they saw fit, for example with one
296 participant claiming more than the other, or with both participants claiming equally many. They
297 were also instructed that their pair should attempt to obtain as many points as possible. There was
298 no communication between the individuals beyond being able to see which items were claimed by
299 themselves, and which by their collaborator.

300 Pairs were constructed by randomly assigning two participants to the same game. Each pair
301 completed 25 trials, after which each individual indicated on a visual analogue scale how good they
302 thought the collaboration was, how fair they thought their collaborator was, and how likeable they

303 thought their collaborator was. Each individual took part in two or three consecutive collaborative
304 sessions.

305

306 **Performance-dependent group reward**

307 Choosing the appropriate reward scheme for the joint visual short-term memory task was
308 not a trivial matter. Ideally, individuals were rewarded as a function of their performance. Here,
309 performance (recall error) was defined as the absolute circular difference between the prompted
310 target's orientation and the associated response. This error existed in a space between 0 degrees
311 (perfect recall) and 90 degrees (the largest possible circular distance between target and response).
312 The most straightforward scheme would entail a linear relationship between individuals' recall error
313 and their reward. However, recall error is rarely uniformly distributed: Participants are generally
314 more likely to respond around 0 error, with the exact shape of the distribution depending on their
315 visual short-term memory capacity κ . A non-linear reward scheme, such as a normal distribution,
316 could address this.

317 We anticipated three ways to play the joint visual short-term memory game outlined above.
318 The first option was an **equal** division of labour, with each participant claiming half of the items to
319 remember.

320 The second option was a **proportional** division of labour, with each participant claiming a
321 proportion of the items that corresponded to the ratio of participants' visual short-term memory
322 capacities. For example, if one participant's capacity κ was 4 bits and the other participant's was 2.4
323 bits, a proportional division of 8 items would have had the higher-capacity participant claim 5 and
324 the lower-capacity participant claim 3.

325 Finally, there was a **paradoxical** division of labour, which could have been optimal
326 depending on how strongly the reward scheme penalised higher errors. Specifically, if low errors

327 were to receive disproportionately larger rewards, the optimal strategy would have been to have the
328 lower-capacity player claim all but one item. The higher-capacity individual would have then
329 claimed only one item, and would thus have achieved a very high reward.

330 To illustrate the effect of different reward functions, we simulated the expected joint rewards
331 between two players of different visual short-term memory capacities κ who divided labour equally,
332 proportionally, or paradoxically (**Figure 3**).

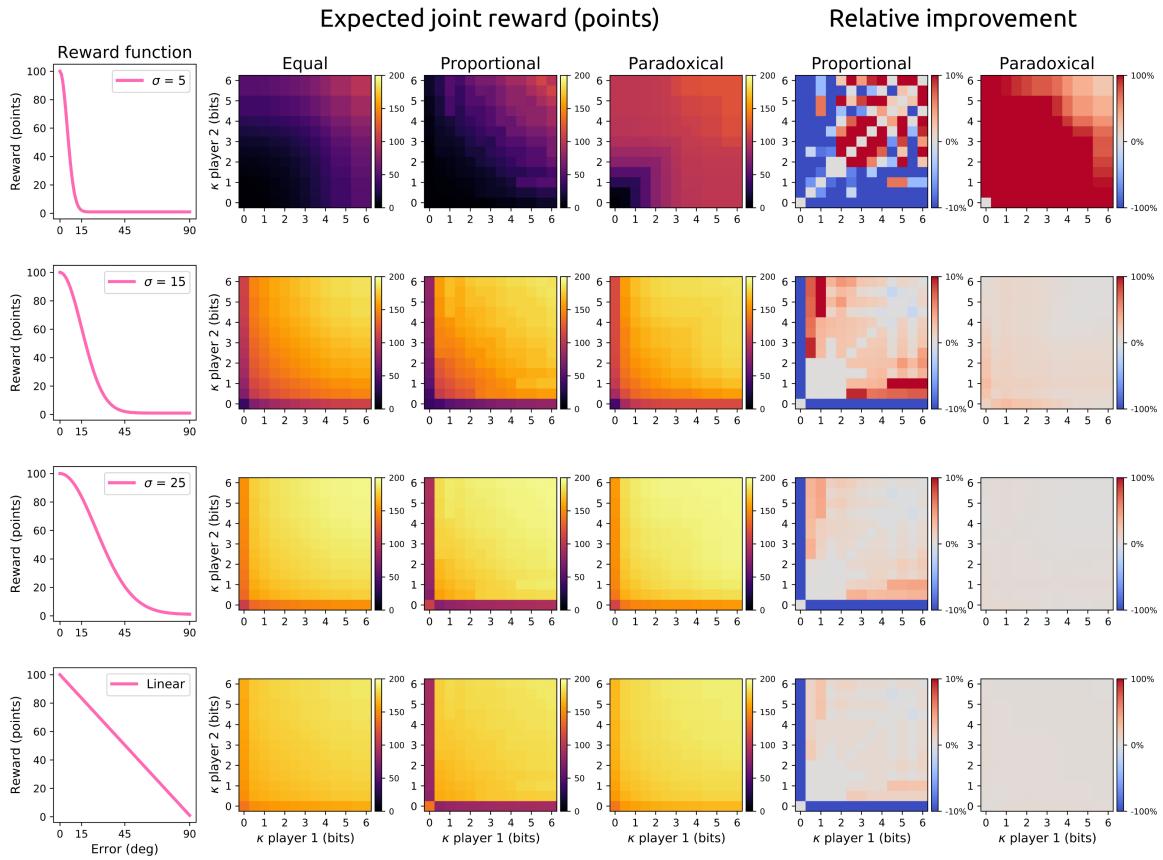
333 When only very low errors would receive a high reward (normal distribution with a low
334 standard deviation), a paradoxical division of labour would be optimal, whereas it would be hard to
335 predict when proportional division would work in a pair's favour (**Figure 3**, top row).

336 When higher errors would be punished less severely (e.g. via a normal distribution with a
337 high standard deviation, or a linear function), paradoxical division of labour would become less
338 ideal. However, although optimal, proportional division would give only a small benefit (**Figure 3**,
339 bottom two rows).

340 The ideal reward function would incentivise proportional division of labour while not
341 making paradoxical division of labour too tempting. Our simulations indicated that this ideal reward
342 function was a normal distribution with a standard deviation that resulted in the incentivisation of
343 low errors without penalising higher errors too strongly (**Figure 3**, second row).

344 Because of the outlined simulations, we opted for defining our individual reward function as
345 a normal distribution with a standard deviation of 17 degrees (centred around 0 degrees). This
346 corresponded with the average standard deviation of errors in a pilot experiment using coloured
347 stimuli (Dalmaijer, Poullias, Somai, & Husain, 2017).

348



349

Figure 3 – Different reward schemes and their expected outcomes for players of different short-term memory capacities. The leftmost column illustrates the reward function, with error (absolute difference in degrees of response and stimulus orientation) on the x-axis, and reward on the y-axis. The second, third, and fourth column show the expected reward under three different strategies. In the equal strategy, collaborators divide items equally. In the proportional strategy, participants divide items proportionally to their visual short-term memory capacity. In the paradoxical strategy, the lower-capacity player takes on all but one of the items, and the higher-capacity player takes on only one item. The x-axis denotes the visual short-term memory capacity of the first player, and the y-axis that of the second player. The colouring indicates the expected group reward, with brighter colours indicating higher rewards. Finally, the penultimate and rightmost columns indicate the relative improvement of the proportional and paradoxical strategies compared to equal division. These columns show that steep reward functions (disproportionality rewarding lower errors) result in high expected benefits from the paradoxical strategy, whereas less steep reward functions do not. However, less steep reward functions also decrease the expected benefit from using a proportional strategy. The optimal reward function, which incentivises the proportional strategy but not the paradoxical, seems to lie in the middle.

350

351 Statistical analyses

352 We considered each of the 32 unique pairs an independent measurement. We divided the
353 individuals within each pair as “higher capacity” and “lower capacity”, according to their visual
354 short-term memory capacity that was independently determined through the individual assessment.

355 Our main hypothesis was that within each pair, the higher-capacity individual would claim
356 more items than the lower-capacity individual. To test this, we used an independent-samples t-test to
357 test whether the average number of claimed items was different between the higher and lower-
358 capacity individuals. We also hypothesised the difference in claimed items would correlate with
359 how participants rated the collaboration, and thus used a linear regression to verify whether the
360 difference in claimed items (relative to the difference in capacity) predicted collaboration
361 judgement of both participants within a pair.

362 Earlier research indicated that in joint visual search, two collaborators divide items in search
363 space, for example by dividing the field in a top and a bottom half (Brennan et al., 2008; Niehorster
364 et al., submitted). To assess whether this pattern also emerged in our data, we computed the average
365 location of all items claimed by each individual. We then computed the average distance, and the
366 standard deviation of the distance (across trials) between each individuals’ average item location.
367 The average unsigned distance reflects the extent to which individuals divided the space, and the
368 standard deviation reflects how consistent they were in dividing the space in the same way. We
369 correlated these two outcome measures with the ratings each player assigned the collaboration
370 quality, and the fairness and likeability of their collaborator.

371 Finally, to assess the effects of personality traits on division of labour, we computed the
372 differences in personality questionnaire scores between the higher-capacity and lower-capacity
373 individuals within each pair. We then used a multivariable linear regression with the difference in

374 capacity as co-variate, as well as the differences in the subscales of the Apathy Motivation Index
375 (Ang et al., 2017) and Big Five personality traits (Gosling et al., 2003) questionnaires.

376 Because of our relatively low number of samples, we used Kendall's rank correlation
377 coefficient τ (Kendall, 1938). This coefficient requires fewer samples than Spearman's and
378 Pearson's correlation coefficients to obtain the same confidence interval (Bonett & Wright, 2000).
379 Specifically, to detect a strong effect (Pearson $R > 0.8$, equivalent to Kendall $\tau > 0.6$) with a Fisher
380 confidence interval width of 0.3 at an α level of 0.05, one would need a sample of 35 (Bonett &
381 Wright, 2000).

382 Note that our sample size was limited primarily due to practical limitations: We depended on
383 the likelihood of a large number of participants to be available at the same time. As a result, we
384 were left with sufficient power to detect strong effects, but not necessarily to detect weaker effects.
385 In addition, we required quite a few comparisons to address our hypotheses. We thus decided on the
386 rather conservative Bonferroni correction for multiple comparisons (Dunn, 1961). Specifically, we
387 set the statistical significance threshold α to 0.05, and performed 23 tests, plus a multivariable
388 regression with 9 predictors (which includes 1 F -test, and 10 t -tests). We thus set our corrected
389 significance threshold to 0.0015 (equivalent to a correction of 34 independent tests). All p values at
390 or below this level will be considered statistically significant.

391

392 Open methods and data

393 Both experiments were programmed in Python (Dalmaijer, 2017; Van Rossum & Drake,
394 2011), using the PyGaze (Dalmaijer, Mathôt, & Van der Stigchel, 2014) and PsychoPy (Peirce,
395 2007) packages. Data extraction, analysis, and visualisation were also performed in Python, using
396 the NumPy and SciPy libraries (Oliphant, 2007), and the Matplotlib package (Hunter, 2007).

397 Visualisation colours were selected from the colourblind-friendly ‘points of view’ palette (Wong,
398 2011). The multivariable regression was performed in JASP, version 0.9.0.1 (JASP Team, 2016).

399 All code and data has been made available online, in a repository hosted on GitHub:
400 https://github.com/esdalmaijer/2017_Lund_Collaborative_short-term_memory_game.

401

402

403 Results

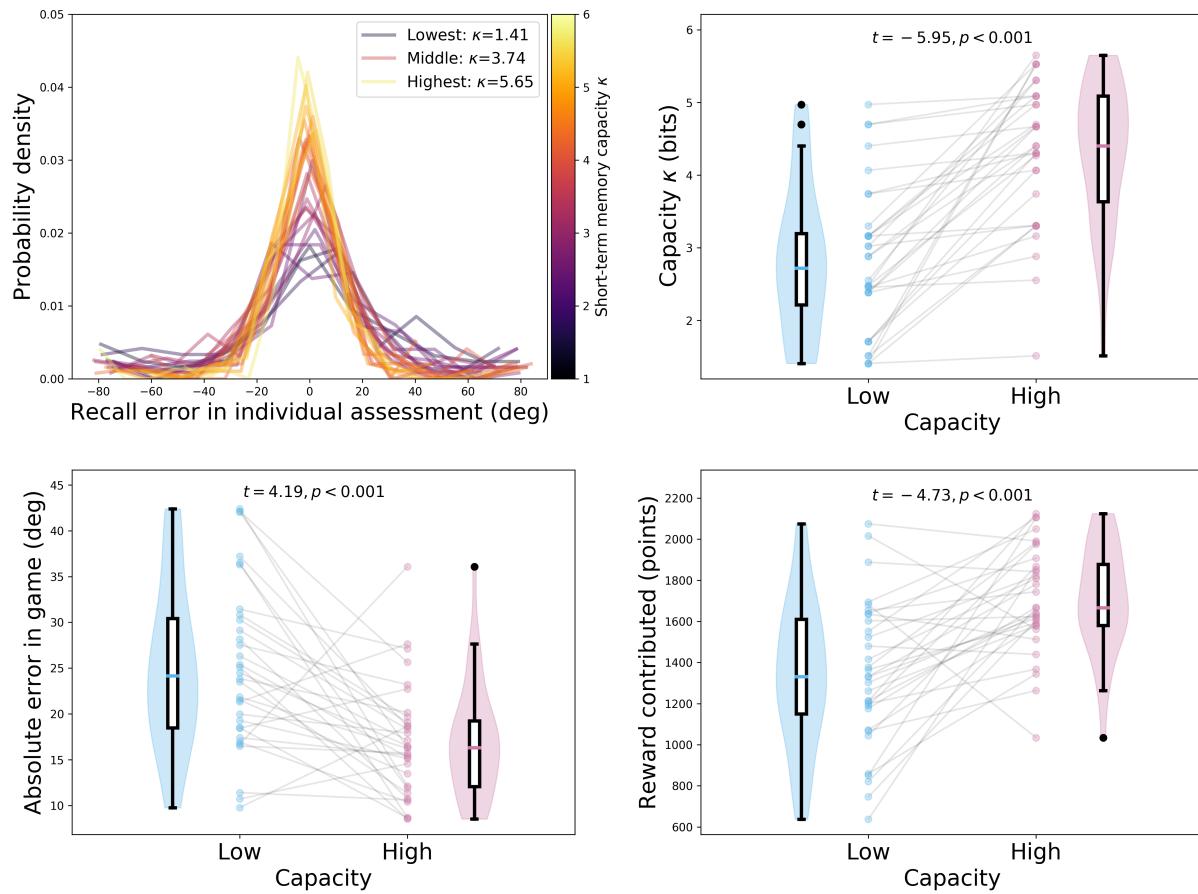
404

405 Manipulation checks

406 There was a good range in the distribution in errors and associated visual short-term
407 memory capacities computed from the individual assessments (**Figure 4A**). The associated visual
408 short-term memory capacities ranged from $\kappa=1.41$ to $\kappa=5.65$, with a mean of 3.62, and a standard
409 deviation of 1.22 bits. This means our sample showed reasonable variance in short-term memory
410 capacity, in line with population estimates (Luck & Vogel, 2013).

411 In the joint visual short-term memory task, participants were paired up randomly to form
412 pairs, and our analyses focussed on whether within a pair the lower-capacity individuals would
413 differ from higher-capacity individuals (for the spread in capacity differences, see **Figure 4A-B**).
414 Indeed, the lower-capacity individuals showed higher recall errors than their higher-capacity
415 collaborators [$t=-4.19, p<0.001$] (**Figure 4C**). Consequently, they contributed significantly fewer
416 points to the group total [$t=-4.73, p<0.001$] (**Figure 4D**).

417



418

A) Histograms showing the probability density of errors in the individual visual short-term memory assessment of all participants. The narrower the curve, the higher the associated short-term memory capacity. **B)** When pairs are sorted according to relatively differences in short-term memory capacity, the lower-capacity and higher-capacity individuals show a large variety in capacity differences. **C)** Across all pairs, lower-capacity individuals (blue) produce higher average absolute errors than higher-capacity individuals (pink). **D)** Across all pairs, lower-capacity individuals (blue) contribute fewer points to the group reward than higher-capacity individuals (pink). Plots show individual samples as coloured dots, observation density in a violin plot, and group median and quartile ranges (plus ‘outliers’ as black dots) as box-plots nested in the violin plots. Observations from the same pair are connected by grey lines.

419

420 No capacity-dependent division of labour

421 Our hypothesis was that players would divide the number of items between themselves

422 according to their capacities. However, there was no significant difference in how many items were

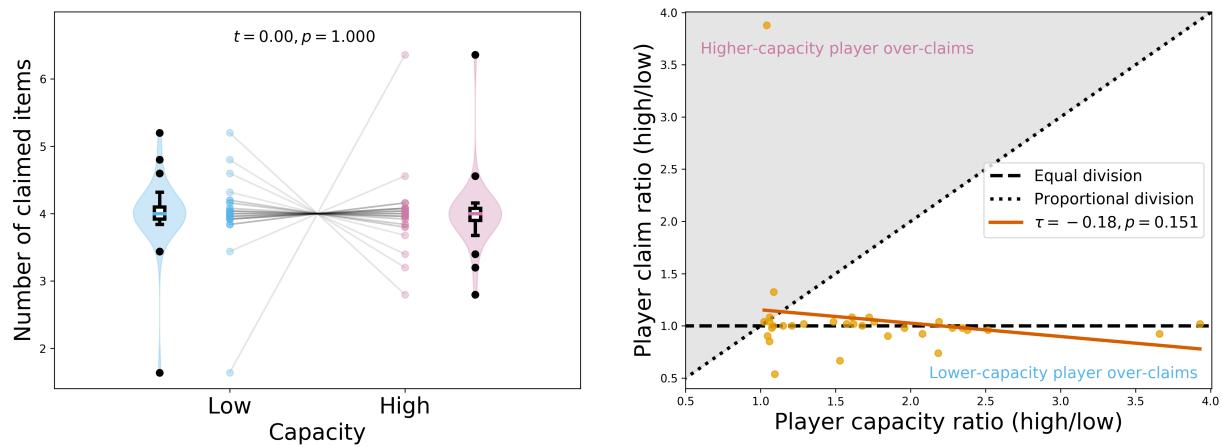
423 claimed by the lower and higher-capacity individuals in the joint visual short-term memory game,

424 [$t=0.00, p=1.000$] (Figure 5A).

425 In addition, we computed the capacity ratios of all pairs, where higher numbers indicated a
 426 larger discrepancy between the players. We also computed the ratio in the number of items each
 427 player claimed, where higher numbers again indicated a larger discrepancy between players. If our
 428 hypothesis were true, there should have been a significant correlation between the players' capacity
 429 ratio and their claim ratio. This correlation was not observed [$\tau=-0.18, p=0.151$], and in fact nearly
 430 all claim ratios were instead close to 1 (**Figure 5B**).

431 These results indicate that participants divided labour equally rather than proportionally.

432



433

Figure 5 – A) When pairs are sorted according to relatively differences in short-term memory capacity, the lower-capacity and higher-capacity individuals do not differ in how many items each claims. Individual samples are shown as coloured dots, observation density in a violin plot, and group median and quartile ranges (plus 'outliers' as black dots) as box-plots nested in the violin plots. Observations from the same pair are connected by grey lines. **B)** There is no correlation between the ratio of visual short-term memory capacity within each pair, and the ratio between the number of claimed items within each pair. In fact, the vast majority of claim ratios lies around one, indicating most pairs divided items roughly equally.

434

435 Division of labour does not affect collaboration perception

436 To assess whether the division of labour affected collaboration perception, we computed the
 437 discrepancy between expected (based on capacity ratios) and actual (based on claim ratios) division

438 of labour. This was simply the signed difference between the claim ratio and the capacity ratio
439 (claim minus capacity), where positive values indicated that the higher-capacity player within a pair
440 claimed more items than they should have (if players divided items proportionally to their
441 capacities), and negative values that the lower-capacity player over-claimed.

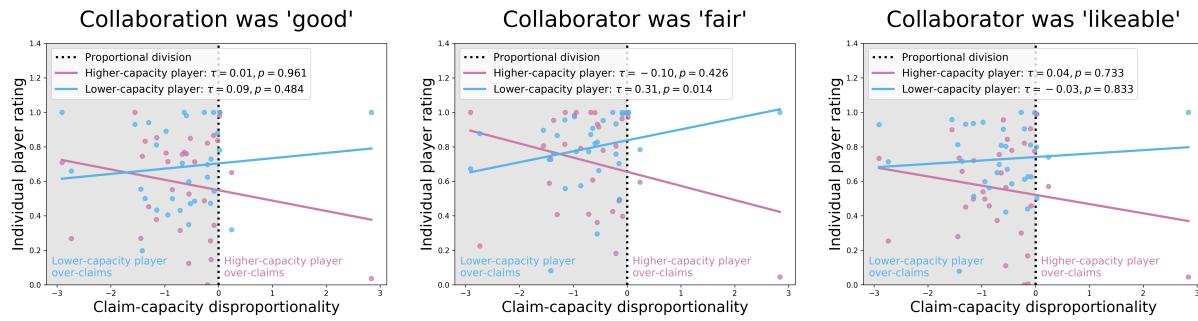
442 We hypothesised that players would have rated the collaboration better when they and their
443 collaborator had shared items proportionally. Hence, we expected a correlation between the
444 difference in claim and capacity ratios on the one hand, and the average collaboration rating on the
445 other. However, there was no statistically significant relationship between the claim-capacity
446 disproportionality and the rating of how good the collaboration was given by higher-capacity
447 individuals [$\tau=0.01, p=0.961$], nor with that given by lower-capacity individuals [$\tau=0.09, p=0.484$]
448 (**Figure 6A**). The same was true for the ratings given by higher-capacity [$\tau=-0.10, p=0.426$] and
449 lower-capacity individuals' [$\tau=0.31, p=0.014$, not significant at $\alpha=0.0015$] of how fair the other
450 player was (**Figure 6B**), and for the ratings given by higher-capacity [$\tau=0.04, p=0.733$] and lower-
451 capacity individuals' [$\tau=-0.03, p=0.833$] of how likeable the other player was (**Figure 6C**).

452 These results indicate that participants did not factor in proportional division of labour in
453 their appreciation of the quality of their collaboration, or the fairness or likeability of their
454 collaborator.

455

456

457



458

Figure 6 – Correlations between the disproportionality between the visual short-term memory capacity and claimed items ratios within each pair. Negative disproportionality values (x-axis) indicate the lower-capacity player (blue) claimed more items than they perhaps should have given their short-term memory capacity, whereas positive values indicate the higher-capacity player (pink) claimed more. A disproportionality of 0 constitutes perfectly proportional division of labour. There was no correlation between disproportionality and ratings of collaboration quality (A), collaborator fairness (B), and collaborator likeability (C) given by either of the participants.

459

460 Collaboration perception was driven by behavioural volatility

461 We hypothesised that division of items according to their position on the screen would affect
 462 collaboration perception. Specifically, participants in joint visual search have been shown to
 463 regularly divide the item space in two halves, even when their only means of communication was
 464 real-time feedback on where the other participant was looking (Brennan et al., 2008). Because the
 465 spatial structure and feedback matches this type of experiment, we hypothesised that adherence to
 466 the same type behaviour would have positively affected collaboration perception.

467 To verify this, we computed the average location of items claimed within each trial by each
 468 participant within a pair. We then computed the average distance between these locations within
 469 each pair, where higher values indicated that participants avoided each other. In addition, we
 470 computed the standard deviation of the distance between the players within each pair, where higher
 471 values indicated less consistent behaviour.

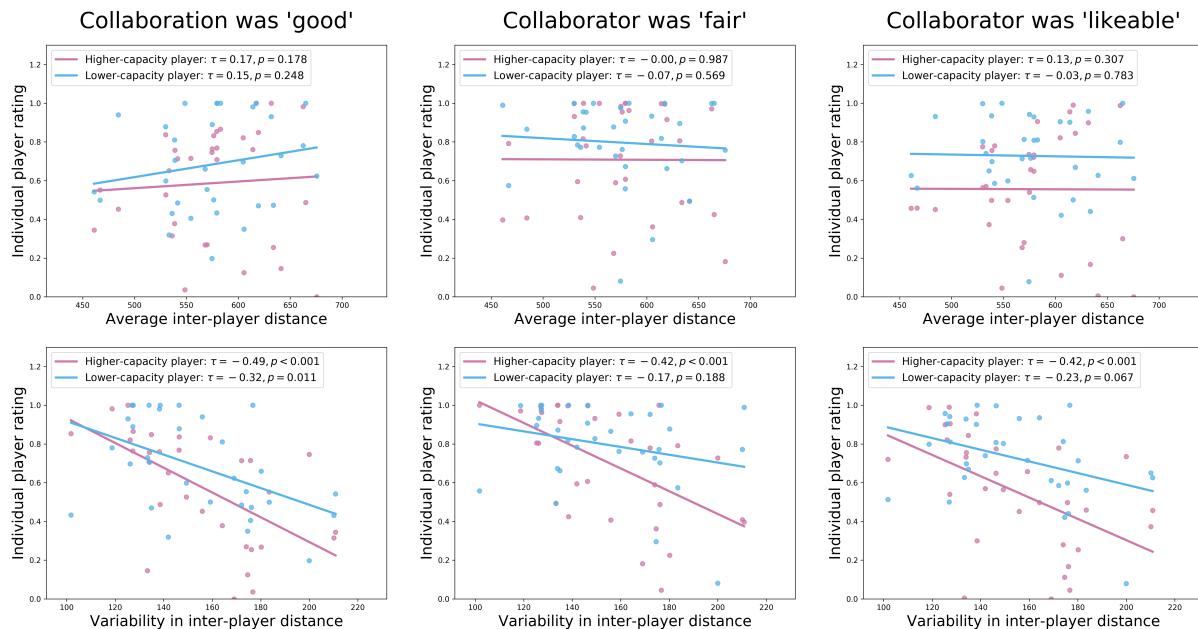
472 There were no statistically significant correlations between the average inter-player distance
 473 and collaboration ratings given by higher-capacity [$\tau=0.17, p=0.178$] or lower-capacity individuals

474 [$\tau=0.15, p=0.248$], or with the collaborator fairness ratings given by higher-capacity [$\tau=0.00,$
475 $p=0.987$] or lower-capacity individuals [$\tau=-0.07, p=0.569$], or with the collaborator likeability
476 ratings given by higher-capacity [$\tau=0.13, p=0.307$] or lower-capacity individuals [$\tau=-0.03, p=0.783$]
477 (**Figure 7**, top row).

478 There were statistically significant relationships between the standard deviation of the inter-
479 player distance and ratings given by higher-capacity individuals on good collaboration [$\tau=-0.49,$
480 $p<0.001$], collaborator fairness [$\tau=-0.42, p<0.001$], and collaborator likeability [$\tau=-0.42, p<0.001$].
481 Although lower-capacity players showed the same pattern, the relationship between the standard
482 deviation of the inter-player distance and ratings given by lower-capacity individuals on good
483 collaboration [$\tau=-0.32, p=0.011$], collaborator fairness [$\tau=-0.17, p=0.188$], and collaborator
484 likeability [$\tau=-0.23, p=0.067$] were not statistically significant at our corrected α level of 0.0015. It
485 should be noted that the distance variability did not correlate with task performance as measured by
486 joint reward [$\tau=-0.12, p=0.347$] (**Figure 7**, bottom row).

487 These results indicate that at least the higher-capacity players rated the quality of
488 collaboration, as well as the fairness and likeability of their collaborator, lower as a function of the
489 variability of the distance between each collaborators' claimed items. This was true despite this
490 variability not affecting group performance.

491



492

Figure 7 – Top row: The distance between the average location of items claimed by the higher-capacity (pink) and lower-capacity individual (blue) within each pair did not correlate with individuals' ratings of how good the collaboration was (A), how fair the collaborator was (B), or how likeable the collaborator was (C). **Bottom row:** The variability in distances between the average location of items claimed by the higher-capacity (pink) and lower-capacity individual (blue) within each pair did correlate with individuals' ratings of how good the collaboration was (D), how fair the collaborator was (E), or how likeable the collaborator was (F); at least for those given by higher-capacity players. (Correlations for ratings given by lower-capacity players were in the same direction, but not statistically significant at our corrected α level of 0.0015.)

493

494 Apathy and Big Five traits affect number of claimed items

495 A secondary objective of our study was to assess the effects of apathy as measured by the
 496 Apathy Motivation Index (Ang et al., 2017) and Big Five personality traits as measured via a brief
 497 questionnaire (Gosling et al., 2003) on the number of items participants claim. Within each pair, we
 498 computed the difference between the higher-capacity and lower-capacity player in the following
 499 independently assessed traits: visual short-term capacity (measured in our independent assessment);
 500 behavioural activation apathy, social motivation apathy, and emotional stability (Apathy Motivation
 501 Index subscales); extraversion, agreeableness, conscientiousness, emotional stability, and openness
 502 (Big Five subscales). We then used these nine differences as co-variates in a linear multivariable

503 regression to predict the differences in the average number of claimed items between the higher and
504 lower capacity players within each pair.

505 The total difference in the number of claimed items was statistically significantly predicted
506 by the complete model [$F(9,22) = 5.81, p < 0.001$]. Significant predictors of the difference in the
507 number of claimed items were social motivation apathy [$t=4.116, p < 0.001$], and conscientiousness
508 [$t=-3.67, p=0.001$]. These results indicate that pairs with a higher difference in social motivation
509 and conscientiousness also showed a higher difference in the number of claimed items.

510 Non-significant predictors at the α level of 0.0015 were the differences in visual short-term
511 memory capacity [$t=1.84, p=0.079$], behavioural activation apathy [$t=-1.86, p=0.077$], emotional
512 stability [$t=1.60, p=0.124$], openness [$t=2.19, p=0.040$], extraversion [$t=1.58, p=0.130$], and
513 agreeableness [$t=1.49, p=0.150$]. Emotional stability as measured with a short-form Big Five
514 questionnaire did not reach statistical significance at the α level of 0.0015 [$t=3.49, p=0.002$].

515 Although higher differences in social motivation apathy work in the opposite direction to
516 higher differences in conscientiousness, it is hard to establish the exact origin of these effects. This
517 is because they were computed using difference scores, and are dependent on the co-variation of
518 other variables.

519

520

521 Discussion

522

523 In a new collaborative visual short-term memory task, we leveraged individual differences
524 in cognitive ability, and employed a non-linear performance reward structure to incentivise
525 collaboration between randomly assigned pairs of participants. Paired individuals did not know who

526 their collaborator was, and could not communicate verbally, but they did receive real-time feedback
527 on each other's behaviour. Our results indicate that collaborators did not share their cognitive
528 workload proportionally to their individual capacities, but instead divided labour equally.
529 Furthermore, the lack of adherence to proportional division did not affect individuals' ratings of the
530 collaboration's quality, nor their collaborator's fairness or likeability. Instead, collaboration ratings
531 correlated strongly with the variability in the distance between locations of items claimed by each
532 individual. Interestingly, this variability did not correlate with individuals' performance, and was
533 thus irrelevant to the task. Finally, our results suggest that individual differences in social apathy
534 and conscientiousness correlated with differences in the proportion of labour within a pair.

535 Others have recently demonstrate the existence of an equality bias among collaborators who
536 had to make a shared perceptual decision (Mahmoodi et al., 2015). That work suggested pairs of
537 collaborators value each other's opinion equally, even when one collaborator was markedly better
538 (whether naturally, or through experimenter manipulation). What remained unclear was whether a
539 discrepancy in collaborators' ability would prompt the higher-ability individual to help the lower-
540 ability individual.

541 Our task differed from Mahmoodi and colleagues' study in that it allowed collaborators to
542 divide labour amongst themselves. Because they received feedback during a series of trials,
543 collaborators were given the opportunity to assess the difference in ability between them.
544 Furthermore, they were subject to a non-linear group reward scheme that incentivised not only
545 individuals' own performance, but also that of their collaborator. Hence, in our experiment
546 collaborators had information on each other's ability, were able to act on their ability discrepancy
547 by proportionally dividing labour, and also incentivised to do so to maximise reward. However, the
548 vast majority of collaborators opted for dividing items equally, which suggests that the equality bias
549 in opinions demonstrated by (Mahmoodi et al., 2015) extends to the division of labour.

550 Another interesting finding we highlight here is that individuals' rating of the quality of
551 collaboration was unaffected by whether or not labour was shared equally. In fact, particularly the
552 higher-ability individual within each pair instead rated the collaboration quality less high as a
553 function of the variability in their pair's behaviour. Specifically, if the distance between their own
554 and their collaborator's average location of claimed items varied a lot between trials, they were
555 more likely to rate the collaboration as less good, and also to rate their collaborator as less fair and
556 less likeable. This is despite the fact this distance variability did not correlate with performance, and
557 was thus unrelated to reward.

558 Earlier work suggests that participants in collaborative cognitive tasks prefer to divide their
559 workspace in halves (Brennan & Enns, 2015; Brennan et al., 2008; Neider et al., 2010), and
560 minimise overlap in self-assigned workspace (Niehorster et al., submitted). This maximises the
561 distance between collaborators, perhaps so they avoid getting in each other's way. Our work
562 suggests that it is not the distance between collaborators, but instead the stability in distance that
563 determines the perceived collaboration quality. Hence, behavioural stability is a trait that individuals
564 seem to prefer in their collaborators.

565 Finally, the results presented here indicate that inter-collaborator differences in cognitive
566 ability (visual short-term memory capacity) did not drive labour division, but that differences in
567 personality traits do. Specifically, inter-collaborator differences in lack of social motivation, or
568 'social apathy' (Ang et al., 2017), correlated with differences in the number of items each claimed.
569 The same is true for the Big Five personality trait conscientiousness, albeit in the opposite direction.
570 Note that these findings should be interpreted more carefully, as they originate from difference
571 scores (which obscure directionality), and because they originate from a model with a large number
572 of co-variates.

573 We should note that the research presented here is highly exploratory, and that our sample of
574 32 pairs only provided enough statistical power to detect large effects. We took this into account in
575 our statistical analyses by being conservative and applying a stringent Bonferroni correction,
576 resulting in a statistical significance threshold of 0.0015. Although we are reasonably confident that
577 the behavioural effects we report are robust, we would like to stress the possibility of smaller effects
578 having gone undetected. We have published our methods and data in a public repository, and invite
579 other researchers to use these tools to replicate and extend our findings.

580

581 Conclusion

582 In conclusion, we explored collaborative behaviour in a novel joint visual short-term
583 memory task, and found that collaborators divide cognitive labour equally between them. This was
584 true even when there was a large discrepancy between collaborators' (independently assessed)
585 short-term memory capacities, and when dividing labour proportionally would have yielded higher
586 group rewards. Paradoxically, participants did not weigh proportional division of labour into their
587 rating of collaboration quality. Instead, high collaboration ratings were associated with low
588 behavioural variability, despite this not relating to group performance.

589 We conclude that in a collaborative task with clear performance feedback but without verbal
590 communication, individuals prefer to divide labour equally and in a consistent pattern, even when
591 proportional division would be optimal.

592

593

References

594

- Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, 97, 79–97. <https://doi.org/10.1016/j.cogpsych.2017.07.001>
- Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. *Psychological Science*, 15(2), 106–111. <https://doi.org/10.1111/j.0963-7214.2004.01502006.x>
- Ang, Y.-S., Lockwood, P., Apps, M. A. J., Muhammed, K., & Husain, M. (2017). Distinct Subtypes of Apathy Revealed by the Apathy Motivation Index. *PLOS ONE*, 12(1), e0169938. <https://doi.org/10.1371/journal.pone.0169938>
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual Working Memory Represents a Fixed Number of Items Regardless of Complexity. *Psychological Science*, 18(7), 622–628. <https://doi.org/10.1111/j.1467-9280.2007.01949.x>
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. In *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Bays, P. M. (2015). Spikes not slots: noise in neural populations limits working memory. *Trends in Cognitive Sciences*, 19(8), 431–438. <https://doi.org/10.1016/j.tics.2015.06.004>
- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10), 1–11. <https://doi.org/10.1167/9.10.7>
- Bays, P. M., & Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science*, 321(5890), 851–854. <https://doi.org/10.1126/science.1158023>
- Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall, and Spearman correlations. *Psychometrika*, 65(1), 23–28.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, 11(5), 1–34. <https://doi.org/10.1167/11.5.4>

- Brennan, A. A., & Enns, J. T. (2015). When two heads are better than one: Interactive versus independent benefits of collaborative cognition. *Psychonomic Bulletin & Review*, 22(4), 1076–1082. <https://doi.org/10.3758/s13423-014-0765-4>
- Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3), 1465–1477. <https://doi.org/10.1016/j.cognition.2007.05.012>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.
- Dalmaijer, E. S. (2017). *Python for experimental psychologists*. Abingdon, Oxon; New York, NY: Routledge.
- Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, 46(4), 913–921. <https://doi.org/10.3758/s13428-013-0422-2>
- Dalmaijer, E. S., Poullias, C., Somai, R., & Husain, M. (2017). When is reward-associated information prioritised in visual working memory? In *Journal of Vision* (Vol. 17(10), p. 869). St. Pete Beach, FL: The Association for Research in Vision and Ophthalmology.
- Dunn, O. J. (1961). Multiple Comparisons among Means. *Journal of the American Statistical Association*, 56(293), 52–64. <https://doi.org/10.1080/01621459.1961.10482090>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- JASP Team. (2016). JASP (Version 0.7.1.12). Retrieved from <https://jasp-stats.org>
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1–2), 81–93. <https://doi.org/10.1093/biomet/30.1-2.81>

- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. <https://doi.org/10.1214/aoms/1177729694>
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, 17(8), 391–400. <https://doi.org/10.1016/j.tics.2013.06.006>
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347–356. <https://doi.org/10.1038/nn.3655>
- Mahmoodi, A., Bang, D., Olsen, K., Zhao, Y. A., Shi, Z., Broberg, K., ... Bahrami, B. (2015). Equality bias impairs collective decision-making across cultures. *Proceedings of the National Academy of Sciences*, 201421692. <https://doi.org/10.1073/pnas.1421692112>
- Messmer, N., Leggett, N., Prince, M., & McCarley, J. S. (2017). Gaze Linking in Visual Search: A Help or a Hindrance? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1376–1379. <https://doi.org/10.1177/1541931213601828>
- Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, 17(5), 718–724. <https://doi.org/10.3758/PBR.17.5.718>
- Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., & Hooge, I. T. C. (submitted). Searching with and against each other: spatiotemporal coordination of visual search behavior in collaborative and competitive settings.
- Nyström, M., Niehorster, D. C., Cornelissen, T., & Garde, H. (2017). Real-time sharing of gaze data between multiple eye trackers—evaluation, tools, and advice. *Behavior Research Methods*, 49(4), 1310–1322. <https://doi.org/10.3758/s13428-016-0806-1>
- Olyphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, 9(3), 10–20. <https://doi.org/10.1109/MCSE.2007.58>

- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Richardson, M. J., Marsh, K. L., & Baron, R. M. (2007). Judging and actualizing intrapersonal and interpersonal affordances. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 845–859. <https://doi.org/10.1037/0096-1523.33.4.845>
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–76. <https://doi.org/10.1016/j.tics.2005.12.009>
- Van Rossum, G., & Drake, F. L. (2011). *Python Language reference manual*. Bristol, UK: Network Theory Ltd.
- Wahn, B., Czeszumski, A., & König, P. (2018). Performance similarities predict collective benefits in dyadic and triadic joint visual search. *PLOS ONE*, 13(1), e0191179. <https://doi.org/10.1371/journal.pone.0191179>
- Wahn, B., Kingstone, A., & König, P. (2018). Group benefits in joint perceptual tasks—a review: Group benefits in joint perceptual tasks. *Annals of the New York Academy of Sciences*. <https://doi.org/10.1111/nyas.13843>
- Wahn, B., Schwandt, J., Krüger, M., Crafa, D., Nunnendorf, V., & König, P. (2016). Multisensory teamwork: using a tactile or an auditory display to exchange gaze information improves performance in joint visual search. *Ergonomics*, 59(6), 781–795. <https://doi.org/10.1080/00140139.2015.1099742>
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 11. <https://doi.org/10.1167/4.12.11>
- Wong, B. (2011). Points of view: Color blindness. *Nature Methods*, 8(6), 441–441. <https://doi.org/10.1038/nmeth.1618>
- Yamani, Y., Neider, M. B., Kramer, A. F., & McCarley, J. S. (2017). Characterizing the efficiency of collaborative visual search with systems factorial technology. *Archives of Scientific Psychology*, 5(1), 1–9. <https://doi.org/10.1037/arc0000030>
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235. <https://doi.org/10.1038/nature06860>