



**HAL**  
open science

## Technical reasoning is important for cumulative technological culture

François Osiurak, Salomé Lasserre, Julie Arbanti, Joël Brogniart, Alexandre Bluet, Jordan Navarro, Emanuelle Reynaud

► **To cite this version:**

François Osiurak, Salomé Lasserre, Julie Arbanti, Joël Brogniart, Alexandre Bluet, et al.. Technical reasoning is important for cumulative technological culture. *Nature Human Behaviour*, 2021, 5 (12), pp.1643-1651. 10.1038/s41562-021-01159-9 . hal-04002136

**HAL Id: hal-04002136**

**<https://hal.science/hal-04002136>**

Submitted on 23 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Technical reasoning is important for cumulative technological  
2 culture

3 François Osiurak<sup>1,2,\*</sup>, Salomé Lasserre<sup>1</sup>, Julie Arbanti<sup>1</sup>, Joël Brogniart<sup>1</sup>,  
4 Alexandre Bluet<sup>1</sup>, Jordan Navarro<sup>1,2</sup>, & Emanuelle Reynaud<sup>1</sup>

5 <sup>1</sup>Laboratoire d'Etude des Mécanismes Cognitifs, Université de Lyon, France

6 <sup>2</sup>Institut Universitaire de France, Paris, France

7 \*Corresponding author (francois.osiurak@univ-lyon2.fr)

8

9 **Abstract**

10 Human technology has evolved in an unparalleled way, allowing us to expand across the  
11 globe. One fascinating question is to understand the cognitive origins of this  
12 phenomenon, which is known as cumulative technological culture (CTC). The dominant  
13 view posits that CTC results from our unique ability to learn from each other. The  
14 cultural niche hypothesis even minimizes the involvement of non-social cognitive skills  
15 in the emergence of CTC, claiming that technologies can be optimized without us  
16 understanding how they work, but simply through the retention of small improvements  
17 over generations. Here we conduct a partial replication of Derex et al.'s (2019)  
18 experimental study and show that the improvement of a physical system over  
19 generations is accompanied by an increased understanding of it. These findings suggest  
20 that technical-reasoning skills (i.e., non-social cognitive skills) are important in the  
21 acquisition/understanding and improvement of technical content (i.e., specific to the  
22 technological form of cumulative culture), thereby making social learning a salient  
23 source of technical inspiration.

24

25 Cumulative technological culture (CTC) refers to the gradual accumulation in the  
26 efficiency and/or complexity of tools/techniques that are too complex to be invented by  
27 a single individual<sup>1-3</sup>. CTC is undoubtedly a social phenomenon in that humans could not  
28 have invented so profuse a range of techniques without social learning. All theories  
29 agree on this point. Instead, much of the debate concerns the nature of the cognitive  
30 skills at the origins of CTC. Many researchers have focused on its social dimension<sup>4,5</sup>,  
31 emphasizing the key role of social cognitive skills such as imitation<sup>6,7</sup>, teaching<sup>8,9</sup> or  
32 theory-of-mind skills<sup>6,10</sup>. In this view, humans are considered as unique social learners<sup>11-</sup>  
33 <sup>13</sup>. The cultural niche hypothesis<sup>1,14,15</sup> even assumes that, in humans, the development of  
34 sophisticated multidimensional technologies (i.e., technologies that combine several  
35 traits) has often occurred without the involvement of specific non-social cognitive  
36 skills/individual-level cognition (e.g., reasoning). In broad terms, “complex technologies  
37 [...] result from the accumulation of many, mostly small, often poorly understood  
38 improvements made across generations linked by cultural transmission. Over time, the  
39 selective retention of improvements gives rise to highly optimized solutions in the  
40 absence of explicit understanding about how these solutions work”<sup>16</sup>. In this respect, a  
41 key prediction of the cultural niche hypothesis is that “the improvement of culturally  
42 evolving technology is not *necessarily* tied to individuals’ level of understanding”<sup>15</sup>.

43 The technical-reasoning hypothesis<sup>17</sup> has recently been developed as an  
44 alternative to account for the cognitive origins of CTC. Its scope is limited to CTC and  
45 does not extend to other forms of cumulative culture (e.g., mathematics, language)<sup>17</sup>,  
46 and diverges in this respect from most of the theories on cumulative culture including  
47 the cultural niche hypothesis. It assumes that the specific human ability to learn from  
48 each other in the technical domain originates precisely in technical-reasoning skills,  
49 namely non-social cognitive skills that make it possible to acquire the understanding of

50 new techniques as well as to potentially improve them. Technical reasoning can be  
51 broadly defined as the ability to reason implicitly about physical object properties. This  
52 non-verbal reasoning is both analogical (i.e., transfer of what is learnt from one situation  
53 to another) and causal (i.e., predicting the effects on the environment)<sup>17</sup>. Technical  
54 reasoning is not synonymous with causal understanding/reasoning. Technical reasoning  
55 is a specific form of causal reasoning directed toward technical content. However,  
56 humans possess other forms of causal reasoning that involve distinct cerebral networks,  
57 such as “semantic/teleological” reasoning (e.g., inferring that eating *Amanita phalloides*  
58 *causes* poisoning) or “theory-of-mind” reasoning (e.g., inferring that the French football  
59 team’s victory *causes* François’ happiness)<sup>17</sup>. Compelling evidence indicates that  
60 technical reasoning is primarily supported by the area PF within the left inferior parietal  
61 lobe, that is, a brain area that is involved when people use and make tools as well as  
62 when they watch others using tools (i.e., a potential situation of social learning)<sup>17-19</sup>. The  
63 technical-reasoning hypothesis stresses the important role of non-social cognitive skills,  
64 and particularly technical reasoning, in the emergence of CTC<sup>17</sup> and, as a result, predicts  
65 that “culturally evolving technology is *necessarily* tied to individuals’ level of  
66 understanding”. In a way, if we envisage a continuum from the idea that technical  
67 reasoning is (a) never to (b) always essential for CTC, the cultural niche hypothesis is  
68 closer to the point (a) whereas the technical-reasoning hypothesis is closer to the point  
69 (b).

70       The recent micro-society study conducted by Derex et al.<sup>16</sup> provided an interesting  
71 experimental design to explore the role of technical reasoning in the specific context of  
72 CTC. The task was to minimize the time it took a wheel to cover 1 m on an inclined track  
73 (Fig. 1a). The wheel had four radial spokes. On each spoke, a weight could be moved  
74 closer or further from the hub of the wheel. The physical system was multidimensional

75 (i.e., it combined at least two traits) because the speed of the wheel depended on the  
76 position of its centre of mass and its moment of inertia. The participants performed the  
77 task as members of chains of 5 participants, each participant being considered as a  
78 “generation” (Fig. 1b). All the participants performed five trials in which they tried to  
79 increase the speed of the wheel by moving the weights along the spokes. They were  
80 provided (except for the first generation) with the last two weight configurations and  
81 the associated speeds of the previous participant in the chain. There were 14 chains of 5  
82 participants in this Configurations condition (i.e., reverse-engineering condition). The  
83 authors also collected data from another 14 chains of 5 participants in a Configurations  
84 + theory condition, in which the participants could generate a written theory about the  
85 wheel system and transmit it to the next participant in the chain (i.e., reverse-  
86 engineering + teaching condition). After the five trials, the participants had to complete  
87 a test measuring their understanding of how the system worked. This test consisted of  
88 10 items. For each item, the participants had to choose which of two wheel  
89 configurations was faster. There were three options: ‘Wheel 1’, ‘Wheel 2’, and ‘No  
90 difference’. Half of the items (i.e., five items) measured the participants’ understanding  
91 of centre of mass and the other half (i.e., five items) their understanding of inertia.

92       The results indicated that the wheel speed became progressively optimized, while  
93 the participants’ understanding of the system did not improve over generations. In  
94 addition, most participants produced unidimensional models related to a specific  
95 dimension (i.e., centre of mass or inertia), which was accompanied by better  
96 understanding of this dimension and worse understanding of the other dimension. This  
97 pattern was explained by the canalizing effect of social transmission on exploration.  
98 Inheriting an inertia-related model prevented the participants from observing the  
99 effects of centre of mass and vice versa, thus giving rise to a compensation phenomenon

100 (i.e., better understanding of one dimension causes worse understanding of the other  
101 dimension). In broad terms, these results supposedly confirm the predictions of the  
102 cultural niche hypothesis. We say “supposedly” because a reexamination of the methods  
103 and the data collected from the understanding test leads us to question the validity of  
104 this conclusion.

105       Firstly, as mentioned above, Derex et al.<sup>16</sup> described a compensation phenomenon  
106 in that better understanding of one dimension was associated with worse  
107 understanding of the other dimension. We do not cast doubt on the canalizing effect of  
108 social transmission on exploration, which is elegantly described by Derex et al.<sup>16</sup>.  
109 However, the question is why better understanding on one dimension was not simply  
110 associated with non-improvement in understanding on the other dimension. An  
111 alternative interpretation is that the negative relationship observed between the centre  
112 of mass and inertia scores did not reflect a compensation phenomenon, but rather the  
113 combination of the canalizing effect described by Derex et al.<sup>16</sup> with a lack of a priori  
114 mutual independence between the two scores because of the introduction of a ‘no  
115 difference’ option. Indeed, a high score on inertia, for instance, implied that the  
116 participants chose the ‘no difference’ option less often for the inertia items. Therefore, if  
117 we consider that the participants could have generated an equiprobable representation  
118 of the answer distribution, having a better understanding of inertia could have led them  
119 to favour ‘no difference’ answers for the less well understood centre of mass dimension,  
120 and vice versa. The corollary is that the greater the score on one dimension, the greater  
121 the number of ‘no difference’ answers on the other dimension. This pattern is precisely  
122 the one observed based on Derex et al.’s<sup>16</sup> data (Extended Data Fig. 1a,b,c; see also Derex  
123 et al.’s Signed Comments). This potential methodological bias also explains why the  
124 participants did not show a global improvement when both scores were pooled

125 together. Note also that the negative relationship between the score on one dimension  
126 and the number of 'no difference' answers on the other dimension indicates that the  
127 participants were relatively good at generating a two-dimensional representation of the  
128 physical system. In sum, this methodological choice (i.e., the introduction of the 'no  
129 difference' option) questions the conclusion that the participants had a "poor  
130 understanding" of the system as argued by Derex et al.<sup>16</sup>.

131 Secondly, and crucially, a major goal of their study was to test whether cultural  
132 adaptations can occur without individual understanding. The best way to test this is to  
133 perform analyses between the wheel speed (e.g., the best speed of the last two trials  
134 with the physical system) and the understanding score. They were not reported in the  
135 study, yet such analyses reveal that the wheel speed is closely tied to individuals' level of  
136 understanding (Extended Data Fig. 1d).

137 Derex et al.'s<sup>16</sup> study presents methodological issues that can explain why they  
138 considered that their results supported the cultural niche hypothesis. A reexamination  
139 of their data questions this conclusion. Nevertheless, this reexamination draws on data  
140 that were collected on the basis of their methodological choices. Therefore, in  
141 Experiment 1, we sought to replicate, at least partly, their study after resolving these  
142 methodological issues to examine once more the role of technical reasoning in CTC. We  
143 did not change the micro-society paradigm (i.e., 14 chains of 5 participants; five trials for  
144 each participant) or the wheel system – although our wheel was faster (Methods). The  
145 main difference concerned the conception of the understanding test (hereafter called  
146 analogous test). We developed a test that consisted of 24 items with four options for  
147 each item (Fig. 1c). The speed difference between the four options of each item was  
148 constant (i.e., an increment of 5 m h<sup>-1</sup>, 10 m h<sup>-1</sup>, or 15 m h<sup>-1</sup>). We also created another  
149 understanding test, which was similar to the aforementioned understanding test except



150 that the configurations represented a wheel with 8 weights, two for each spoke  
151 (hereafter called transfer test; Fig. 1d). As mentioned above, technical reasoning is  
152 analogical, implying that the understanding gained in a technical situation is  
153 spontaneously transferred to other situations. Therefore, in this framework, an increase  
154 in understanding over generations should be found not only for the analogous test but  
155 also for the transfer test. We also collected data from a control group ( $n = 70$ ) who had  
156 no experience with the wheel system in order to obtain a baseline for the two  
157 understanding tests. Finally, we focused only on the Configurations condition, for which  
158 Derex et al.<sup>16</sup> found no improvement in understanding. Indeed, the trend was even  
159 negative). There was therefore less chance of observing an increase in understanding.

160 The results confirmed the predictions of the technical-reasoning hypothesis. The  
161 wheel speed increased over generations (Generation 95% credible interval [CI]: 1.65 to  
162 4.62; median = 3.16 m h<sup>-1</sup>, BF = 42.91; Fig. 2a; for a description of the evolution of wheel  
163 configurations over generations, see Extended Data Fig. 2) in parallel with the  
164 participants' understanding as assessed in both understanding tests (Analogous test:  
165 Generation 95% CI: 0.18 to 4.28; median = 2.14, BF = 24.97; Transfer test: Generation  
166 95% CI: 0.86 to 4.31; median = 2.68, BF = 3.58; Fig. 2b). The control group obtained  
167 understanding scores lower than those of each generation in the experimental group  
168 (Fig. 2b; Supplementary Results Tables 1 and 2). In addition, the estimated Bayes factors  
169 revealed moderate to extreme evidence for the presence of links between the wheel  
170 speed and each of the two understanding scores (Analogous test: Wheel 95% CI: 0.26 to  
171 0.86, median = 0.56, BF = 17.93; Transfer test: Wheel 95% CI: 0.30 to 0.72, median =  
172 0.51, BF > 100; Fig. 2c,d) as well as between the two understanding scores not only in  
173 the experimental group (Analogous score 95% CI: 0.45 to 0.72, median = 0.58, BF > 100;

174 Fig. 2e), but also in the control group ( $\rho_{\text{median}} = 0.58$ ; 95% CI: 0.45 to 0.70,  $\text{BF} > 100$ ;  
175 Fig. 2f).

176 The results of Experiment 1 indicate that the improvement of the wheel system  
177 was accompanied by an increase in understanding at both the group and the individual  
178 level. This increase was observed not only for the analogous test but also for the transfer  
179 test, suggesting that the participants were able to transfer what they understood from  
180 the wheel system to another theoretical wheel system. The addition of the control group  
181 was also informative in that it demonstrated that this experience allowed the  
182 participants in the first generation to outperform controls, who had not experienced the  
183 wheel system at all. This outperformance is difficult to explain other than by an increase  
184 in understanding (i.e., individual-level cognition) due to the experience with the wheel  
185 system. Therefore, even if we had not shown any increase in understanding over  
186 generations as in Derex et al.<sup>16</sup>, it would still be difficult to consider that the participants  
187 in the experimental group had poor understanding of the wheel system, at least when  
188 compared to the control group.

189 Even though these findings provide support for the important role of technical  
190 reasoning in CTC, they do not directly refute the predictions from the cultural niche  
191 hypothesis, in particular because of the presence of significant methodological  
192 limitations (as rightly pointed out in Derex et al.'s Signed Comments). As mentioned, our  
193 wheel was faster than the one used by Derex et al.<sup>16</sup> because the steel axis of our wheel  
194 was heavier than the wooden axis of their wheel (Methods). This resulted in the  
195 detrimental effect that our wheel was less sensitive to the position of the centre of mass  
196 than Derex et al.'s<sup>16</sup> wheel (Extended Data Fig. 3a,b). As a result, our wheel system was  
197 not sufficiently multidimensional, with the result that the participants could have  
198 improved it based primarily on their understanding of a single dimension, namely, the

199 moment of inertia, which is easier to understand. Although the wheel configurations  
200 produced by our participants over generations were generally unbalanced (i.e., a  
201 potential sign of an understanding of centre of mass; Extended Data Fig. 2), the fact  
202 remains that this methodological limitation did not allow us to definitely test the  
203 predictions of the cultural niche hypothesis, which focus on the multidimensional aspect  
204 of technologies. Another limitation concerns the design of our understanding tests, in  
205 which the speed difference between the four options of each item was constant (i.e., an  
206 increment of 5 m h<sup>-1</sup>, 10 m h<sup>-1</sup>, or 15 m h<sup>-1</sup>). However, as also rightly stressed by Derex et  
207 al. (Signed Comments), these differences were obtained by altering the inertia of the  
208 wheel more than the position of its centre of mass. The consequence is that our two  
209 understanding tests could have assessed our participants' understanding of inertia  
210 much more than they did their centre of mass. Taken together, these methodological  
211 limitations could have led us to investigate how the participants improved and  
212 understood a quasi-unidimensional but not a multidimensional physical system.

213         Notwithstanding these limitations, the results of Experiment 1 were promising in  
214 suggesting that an increase of understanding over generations could occur even for a  
215 multidimensional physical system. As stressed by Derex et al.<sup>16</sup>, the participants in their  
216 study had appropriate intuitions about how to maximize acceleration. This is confirmed  
217 by a reexamination of their data, which indicates that the progressive increase in  
218 understanding in the Configurations + theory condition (i.e., the condition for which  
219 they observed a significant increase over generations) mainly concerned the inertia  
220 score and not the centre of mass score (Extended Data Fig. 4). Therefore, even if the  
221 wheel system of Experiment 1 was principally based on the inertia dimension, our  
222 results revealed that individuals can progressively improve their understanding of a  
223 dimension that is initially poorly understood. The corollary is that individuals

224 confronted with a two-dimensional wheel system might be able, over generations, to  
225 combine their progressive understanding of the moment of inertia with their more  
226 intuitive understanding of the position of the centre of mass.

227 For this reason, we conducted a second experiment (Experiment 2), which  
228 overcame the methodological limitations of Experiment 1 and was a more appropriate  
229 way to directly investigate the role of technical reasoning in CTC. As in Experiment 1, we  
230 did not change the micro-society paradigm (i.e., 14 chains of 5 participants; five trials for  
231 each participant) and chose to focus only on the Configurations condition, for which  
232 Derex et al.<sup>16</sup> found no improvement in understanding. We also again used two  
233 understanding tests (i.e., analogous and transfer) and again included a control group,  
234 which did not experience the wheel system. The main modifications concerned the axis  
235 of the wheel, the design of the two understanding tests and the reintroduction of a test  
236 equivalent to the one employed by Derex et al.<sup>16</sup>. Firstly, we modified the axis of our  
237 wheel system in order to obtain a two-dimensional wheel system comparable to the one  
238 used by Derex et al.<sup>16</sup> (Extended Data Fig. 3c,d). Secondly, we modified the two  
239 understanding tests to assess independently the participants' understanding of the  
240 centre of mass and inertia dimensions independently of one another (i.e., 24 items for  
241 each test; 12 centre of mass items and 12 inertia items for each test; four options for  
242 each item). Thirdly, we reintroduced a third understanding test, which was similar to  
243 the one employed by Derex et al.<sup>16</sup> (hereafter called the three-option test; 10 items; five  
244 centre of mass items and five inertia items; three options: 'Wheel 1', 'Wheel 2', and 'No  
245 difference').

246 According to the cultural niche hypothesis<sup>16</sup>, no improvement over generations  
247 should be found for the three understanding tests (i.e., analogous, transfer and three-  
248 option), when the centre of mass and inertia scores are pooled together. This pattern is

249 expected because of the so-called compensation phenomenon, in which an increase in  
250 understanding on one dimension causes a decrease in understanding on the other  
251 dimension. In this respect, a negative link should also be observed between the centre of  
252 mass and inertia scores for the three understanding tests. By contrast, the technical-  
253 reasoning hypothesis predicts an increase in understanding in at least the analogous test  
254 and the transfer test (i.e., when the centre of mass and inertia scores are pooled  
255 together). However, no negative link is expected between the centre of mass and inertia  
256 scores for these two tests. Instead, as explained above, over generations the participants  
257 should combine their progressive understanding of the inertia dimension with their  
258 more intuitive understanding of centre of mass. Concerning the three-option test, the  
259 technical-reasoning hypothesis also predicts an increase in understanding over  
260 generations. However, the methodological limitations of this test (i.e., the introduction of  
261 the 'no difference' option along with the potential equiprobable representation of the  
262 answer distribution; see above) should lead us to reproduce Derex et al.'s<sup>16</sup> finding,  
263 namely, the absence of increase in understanding over generations as well as a negative  
264 link between the centre of mass and inertia scores. Crucially, if this negative link was  
265 reported only for the three-option test but not for the analogous transfer tests, this  
266 would demonstrate that the absence of increase in understanding in Derex et al.'s<sup>16</sup>  
267 study resulted from the methodological choice of using the 'no difference' option in their  
268 understanding test.

269       The results confirmed the predictions of the technical-reasoning hypothesis. The  
270 wheel speed increased over generations (Generation 95% credible interval [CI]: 2.48 to  
271 5.01; median = 3.75 m h<sup>-1</sup>, BF > 100; Fig. 3a; for a description of the evolution of wheel  
272 configurations over generations, see Extended Data Fig. 5) in parallel with the  
273 participants' understanding as assessed in the three understanding tests (Analogous

274 test: Generation 95% CI: 1.37 to 4.78; median = 3.08, BF = 7.88; Transfer test:  
275 Generation 95% CI: 1.15 to 4.19; median = 2.71, BF = 4.79; Three-option test:  
276 Generation 95% CI: 0.02 to 0.64; median = 0.32, BF = 0.39; Fig. 3b,c). The control group  
277 obtained understanding scores lower than those of each generation in the experimental  
278 group (Fig. 3b,c; Supplementary Results Tables 3, 4 and 5). In addition, the estimated  
279 Bayes factors revealed strong to extreme evidence for the presence of links between the  
280 wheel speed and each of the three understanding scores (Analogous test: Wheel 95% CI:  
281 0.51 to 0.92, median = 0.71, BF > 100; Transfer test: Wheel 95% CI: 0.40 to 0.81, median  
282 = 0.61, BF > 100; Three-option test: Wheel 95% CI: 0.05 to 0.13, median = 0.09, BF =  
283 44.88; Fig. 3d,e,h) as well as between the analogous score and the transfer score not  
284 only in the experimental group (Analogous score 95% CI: 0.58 to 0.86, median = 0.72, BF  
285 > 100; Fig. 3f), but also in the control group ( $\rho_{\text{median}} = 0.34$ ; 95% CI: 0.19 to 0.51, BF = >  
286 29.36; Fig. 3g). Links were also generally found between the three-option score and the  
287 analogous and transfer scores in both the experimental group and the control group  
288 (Extended Data Fig. 6).

289       The results for the centre of mass and inertia scores for the three tests also provide  
290 support for the technical-reasoning hypothesis. As predicted, the increase in  
291 understanding for the three tests mainly concerned the inertia score, whereas the centre  
292 of mass score was relatively stable over generations (Extended Data Fig. 7). This finding  
293 is consistent with those of Experiment 1. We also found a negative relationship between  
294 the centre of mass and inertia score for the three-option test, as in Derex et al.<sup>16</sup> (Fig.  
295 4a,b,c). Importantly, this negative relationship was not found for the analogous and  
296 transfer tests (Fig. 4d,e). This strongly suggests that the three-option test is biased by  
297 the introduction of the 'no difference' option and the equiprobable representation of the  
298 answer distribution. This bias, which can be subject to inter-individual variation, can

299 explain why we observed an increase in understanding over generations for the three-  
300 option test, whereas Derex et al.<sup>16</sup> did not. The negative relationship between the centre  
301 of mass and inertia scores in the three-option test observed in Derex et al.<sup>16</sup> as well as in  
302 our experimental group was not reported in the control group (Extended Data Fig. 8).  
303 This corroborates the idea that only the participants in the experimental group were  
304 able to form a correct two-dimensional representation of the wheel system, leading  
305 them to select ‘no difference’ answers for the less understood dimension.

306       Our results are straightforward: What is transmitted from generation to  
307 generation is not the “technique” strictly speaking but the understanding of its  
308 underlying principle. Our study demonstrates that the experience of the wheel system  
309 (i.e., first generation), which was progressively associated with the social transmission  
310 of technical content (i.e., second-to-fifth generations), led the participants in the  
311 experimental group to gradually improve their understanding of how the wheel worked.  
312 Indeed, in both experiments, we observed a substantial increase in wheel speed in the  
313 first generation between the first and fifth trials (see also Derex et al.<sup>16</sup> for a similar  
314 pattern). This on its own points to a significant role of individual-level cognition, which  
315 is corroborated by the substantial increase in understanding among the participants in  
316 the first generation compared to the control group. What is more, in both experiments,  
317 the increase observed for the wheel speed over generations is largely paralleled by the  
318 increase reported for the understanding tests, as is also confirmed by the links reported  
319 between the wheel speed and the understanding scores. It is noteworthy that we  
320 reported this cumulative effect in a Configurations condition, that is, a condition for  
321 which Derex et al.<sup>16</sup> did not report any sign of improvement. Taken together with Derex  
322 et al.’s<sup>16</sup> results (when reexamined), our findings show that non-social cognitive skills,  
323 and particularly technical-reasoning skills, are important in the

324 acquisition/understanding and improvement of technical content – and to avoid  
325 reinventing the wheel. This is also consistent with a growing body of evidence indicating  
326 that learners' *global* technical-reasoning skills are a good predictor of cumulative  
327 performance in micro-society experiments<sup>20-22</sup>. However, the present study  
328 substantially expands these findings by demonstrating that the social transmission of  
329 technical content can also be accompanied by a *specific* improvement of understanding.

330       The cultural niche hypothesis assumes that our reasoning skills can contribute to  
331 the development of sophisticated technologies<sup>14-16,23,24</sup>. Nevertheless, the core  
332 assumption is that these skills do not drive the emergence of CTC because imitation –  
333 the key component for CTC – does not necessarily need such skills. This hypothesis may  
334 suffer from an “under-intellectualization” of the imitation phenomenon. Indeed, the  
335 question of whether nonhuman species can imitate has been the subject of intense  
336 debate for a number of years. For instance, even if chimpanzees can behave in a  
337 conformist manner<sup>25,26</sup>, establish a tradition within a group<sup>27,28</sup>, and rely on cultural  
338 behaviour to address new tool-use problems<sup>29,30</sup>, the question that is still debated is  
339 whether they spontaneously imitate to solve tool-use problems<sup>12,30-34</sup>. In other words,  
340 imitation is not trivial at a cognitive level. Indeed, in acknowledging that causal  
341 reasoning permits the exclusion of irrelevant traits when someone attempts to  
342 reproduce a technology (e.g., the colour of a bow)<sup>14</sup>, the cultural niche hypothesis is to  
343 some extent consistent with this approach. However, it also considers that the copying  
344 of relevant traits (e.g., the plait for attaching the sinew) can be based on the learning of  
345 correlations between these traits and success without requiring the causal  
346 understanding of these connections<sup>14</sup>. The distinction drawn here between learning and  
347 reasoning is a recurrent theme in the literature on animal cognition. The methodology  
348 commonly used is to propose a transfer test to investigate whether the animal has



349 learned *versus* understood the task studied<sup>35</sup>. We employed such a transfer test in  
350 Experiments 1 and 2. We found an increase in understanding over generations for both  
351 experiments, suggesting that the participants did not learn (from the link between  
352 several produced solutions and their success) but gradually improved their  
353 understanding of the wheel system. In addition, it is noteworthy that the participants  
354 could hardly establish a correlation between the relevant traits and success since they  
355 only had access to the two solutions produced by their predecessor and not to several  
356 such solutions and their associated success. In broad terms, these findings provide  
357 support for the idea that, in humans, reproducing a technology made by a predecessor is  
358 accompanied by an increased understanding of it, that is, by an increase of technical-  
359 reasoning skills.

360 In this respect, the technical-reasoning hypothesis offers a comprehensive account  
361 explaining why humans can easily learn technical content from each other: We possess  
362 specific technical-reasoning skills enabling us to acquire the understanding of new  
363 techniques and improve them. This view diverges from the social-centred view, which  
364 attributes the potential uniqueness of human social learning mainly – if not only – to  
365 social cognitive skills. To paraphrase Whiten et al.<sup>34</sup>, the really significant difference  
366 might lie not so much in the social-learning mechanisms themselves but rather in the  
367 complexity of the mechanical actions that humans can represent and, as a result, copy.  
368 This non-social-centred interpretation of imitative skills also opens up new avenues for  
369 explaining why signs of CTC have been reported in New Caledonian crows, a tool-  
370 using/making species that does not engage in complex forms of social learning (e.g.,  
371 imitation, teaching)<sup>36</sup>. This species is nevertheless well known for its remarkable  
372 understanding of the physical world<sup>37</sup>. Interestingly, recent evidence indicates that this  
373 species could even perform reverse engineering<sup>38</sup>, as reported in the present study.

374 In the present study, we chose to put the focus on a reverse-engineering condition,  
375 not because the technical-reasoning hypothesis assumes that such a condition is  
376 sufficient for the emergence of complex forms of CTC, but because it is an interesting  
377 condition to explore the role of technical reasoning in CTC. This indirect social-  
378 transmission condition has the advantage of not involving more elaborated forms of  
379 social learning (e.g., teaching), thereby making it appropriate for investigating more  
380 specifically the non-social cognitive skills at work. Like most theories of CTC or of  
381 cumulative culture more broadly, the technical-reasoning hypothesis acknowledges the  
382 boosting role of more elaborated forms of social learning in CTC<sup>17</sup>. Thus, like the cultural  
383 niche hypothesis, the technical-reasoning hypothesis is consistent with the modern  
384 versions of the Vygotskian intelligence hypothesis, which stress the importance of  
385 collaboration, communication and teaching in the development of cognitive skills<sup>6,8,39</sup>.  
386 However, the distinction between this and other theories of CTC is that the technical-  
387 reasoning hypothesis posits that this boosting role could not have taken place without a  
388 specific cognitive structure (i.e., technical reasoning) which allows us to gradually  
389 increase our understanding of tools/techniques and to potentially show signs of CTC  
390 even with less elaborated forms of social learning (e.g., reverse engineering,  
391 observation)<sup>17</sup>.

392 The technical-reasoning hypothesis also provides a theoretical framework to  
393 overcome the epistemological obstacle that consists in conflating understanding and  
394 explaining<sup>17</sup>. For instance, Derex et al.<sup>16</sup> repeatedly used the term “explicit” to  
395 characterize the concept of causal understanding. This confusion is perpetuated more  
396 generally by the proponents of the cultural niche hypothesis, who have argued that  
397 technologies might have undergone improvement even in preliterate societies. It is true  
398 that human technology has evolved for long periods without individuals being able to

399 produce explicit explanations of the underlying physical principles. Nevertheless, this  
400 does not demonstrate that CTC results from the accumulation of often poorly  
401 understood improvements. After all, humans did not wait for Newton to *understand* that  
402 things can fall. The difference is that Newton attempted to *explain* this phenomenon.  
403 Likewise, Baillargeon and colleagues have shown that infants develop a gradual  
404 understanding of the principle of support<sup>40</sup>. However, most adults find it difficult to  
405 convey this principle explicitly. In other words, understanding and explaining are two  
406 orthogonal cognitive abilities, as stressed by the cognitive niche hypothesis<sup>41,42</sup>, which  
407 distinguishes between know-how (i.e., implicit understanding of the physical world; a  
408 concept close to technical reasoning) and language (explicit explanation). Interestingly,  
409 this discussion opens up avenues that might make it possible to reconcile the cultural  
410 niche hypothesis with the technical-reasoning hypothesis inasmuch as both agree that it  
411 is not necessary to be able to explain the underlying principles of a technique in order to  
412 improve it.

413       To conclude, we would like to stress that our findings are not sufficient to fully  
414 support the technical-reasoning hypothesis. As stated by Derex et al.<sup>16</sup>, the wheel system  
415 is a low-dimensional optimization problem, which provides a conservative test of the  
416 prediction that improvement over generations can occur without an increase in  
417 individuals' understanding. Thus, even if our results invalidate this prediction in this  
418 particular case, evidence is also needed to invalidate it in higher-dimensional  
419 optimization problems. The technical-reasoning hypothesis predicts that, even in such  
420 problems, the improvement should be tied to individuals' level of understanding. Future  
421 work is required to test this possibility. Be that as it may, our findings highlight that the  
422 technical-reasoning hypothesis is a valid alternative to the social-centred view, with the  
423 merit of building a bridge between the classical disciplines interested in cultural

424 evolution (i.e., evolutionary biology, mathematics, anthropology, archaeology)<sup>43</sup> and  
425 cognitive science.

## 426 **Methods**

427 The Ethics Committee of the Lyon Department of Psychology approved the study, and  
428 the procedure was carried out in accordance with the ethical standards of the 1964  
429 Declaration of Helsinki. Informed consent was obtained from all the participants.

430 **Participants.** Two hundred and eighty students at the University of Lyon took part in  
431 the study (Experiment 1:  $M_{\text{age}} = 21.5$ ,  $SD_{\text{age}} = 3.5$ ; 87 women;  $n = 70$  in the experimental  
432 group and  $n = 70$  in the control group; Experiment 2:  $M_{\text{age}} = 20.1$ ,  $SD_{\text{age}} = 2.8$ ; 99 women;  
433  $n = 70$  in the experimental group and  $n = 70$  in the control group). They were randomly  
434 recruited through advertisements posted on social media websites. Contrary to Derex et  
435 al.<sup>16</sup>, the participants were not paid to take part in the experiments.

436 **Experimental apparatus.** The wheel system used in the present study was similar to  
437 that used by Derex et al.<sup>16</sup> (for an illustration, see <https://osf.io/athg5/>; see  
438 Supplementary Methods for more specific details on the wheel system). In Experiment 1,  
439 we reproduced it with the exception of the axis of the wheel. In the original study, it  
440 consisted of a 10.5-cm-long bored-through wooden pole and an 8 mm threaded steel rod  
441 in its centre. In Experiment 1, it took the form of a 10.5-cm-long steel pole. This  
442 difference explains why our wheel was about 80 m h<sup>-1</sup> faster than the one in the original  
443 study. In Experiment 2, the axis of the wheel was the same as in Derex et al.<sup>16</sup>, leading to  
444 a substantial decrease of speed as compared to Experiment 1. Nevertheless, the wheel of  
445 Experiment 2 was still 40 m h<sup>-1</sup> faster than in Derex et al.<sup>16</sup>, due undoubtedly to the fact  
446 that the amount of friction between the wheel and the apparatus was less than in Derex

447 et al.<sup>16</sup>. As described in Extended Data Fig. 3, we conducted a series of pre-tests before  
448 starting Experiment 2 to ensure that the wheel behaved similarly to Derex et al.'s<sup>16</sup>  
449 wheel.

450 **Procedure.** In both experiments, the procedure was similar to that used by Derex et  
451 al.<sup>16</sup>. The experiment took place in an experimental room at the University of Lyon  
452 (around 20 min and 35 min long for the participants in the control group and  
453 experimental group, respectively). The participants in the experimental group sat at a  
454 table placed 2 m from the experimental apparatus. Before the experiment, the  
455 participants completed a consent form. After the experiment, they indicated whether  
456 they had an academic background in engineering or physics.

457 **Experimental design.** In both experiments, the experimental design was also similar to  
458 the one detailed by Derex et al.<sup>16</sup> for the Configurations condition, except that we  
459 developed a new understanding test (i.e., analogous test), created a transfer test, and  
460 recruited a control group as well as used an understanding test similar to the one  
461 employed by Derex et al.<sup>16</sup> (i.e., the three-option test; Experiment 2 only).

462 *Building phase.* This phase was similar in both experiments. Only the participants in the  
463 experimental group were concerned by this phase. Instructions were similar to those of  
464 Derex et al.<sup>16</sup>, except that there was no monetary incentive (see <https://osf.io/athg5/>).  
465 They had 5 trials to optimize the speed of a wheel that descended a 1-m-long inclined  
466 track. They could move 4 weights to any of 12 discrete positions along each spoke and  
467 were free to choose their own configuration (see <https://osf.io/athg5/>). After the  
468 participants confirmed the configuration, the experimenter positioned the weights on  
469 the physical wheel accordingly. The time it took the wheel to travel down the track was  
470 automatically recorded by a computer program (see <https://osf.io/athg5/>). The wheel

471 speed and the associated configuration were then displayed to the participants, who had  
472 as much time as they needed to consult their last two configurations and choose the next  
473 one. Contrary to Derex et al.<sup>16</sup>, we did not use a computer program but a paper-and-  
474 pencil method to display the wheel speeds and the associated configurations (see  
475 <https://osf.io/athg5/>). After three trials, the experimenter reminded the participants  
476 that their last two configurations and the associated speeds would be transmitted to the  
477 next participant in the chain. As in Derex et al.<sup>16</sup>, there were 14 chains of 5 participants  
478 each (i.e., 14 chains in Experiment 1 and 14 chains in Experiment 2). Derex et al.<sup>16</sup>  
479 assigned their participants to sex-segregated chains. Although Derex et al.<sup>16</sup> controlled  
480 for this variable, they did not take it into consideration in their analyses (e.g., as a  
481 random effect), nor did they justify why such a control was needed. This variable can  
482 have an impact in micro-society paradigms, in which participants directly interact  
483 together. However, such an impact is very unlikely here because of the absence of direct  
484 interaction between the participants. The participants were also not informed about  
485 previous participants' identity. For these reasons, we did not assign the participants to  
486 sex-segregated chains contrary to Derex et al.<sup>16</sup>.

487 *Testing phase.* In both experiments, the participants in the experimental group  
488 completed this phase after the building phase. Those in the control group completed  
489 only this phase. Before beginning the test, the participants in the control group were  
490 presented with a picture of the wheel system in order to help them understand the  
491 instructions. The wheel shown in this picture had no weight and was placed in initial  
492 position at the top of the rails. In Experiment 1, all the participants first completed the  
493 analogous test and then the transfer test. They were instructed that they would be  
494 presented with items consisting of 4 wheels and that they would have to guess which of

495 the four wheels would cover the rails faster. They could take as much time as they  
496 needed to complete the items. They had no feedback. All the participants saw the same  
497 items in the same order. Both the analogous and the transfer test consisted of 24 items.  
498 The incremental difference between the four options was 5 m h<sup>-1</sup> (8 items), 10 m h<sup>-1</sup> (8  
499 items), or 15 m h<sup>-1</sup> (8 items). The speed of each option was obtained on the basis of the  
500 physical wheel system. Each item was scored from 1 (the wheel with the lowest speed)  
501 to 4 (the wheel with the highest speed). Therefore, the total possible range of scores for  
502 each understanding test was 24–96. In Experiment 2, all the participants first completed  
503 the three-option test and then the analogous test and the transfer test. For the three-  
504 option test, the participants were instructed that they would be presented with items  
505 consisting of two wheels and that they would have to guess which of these would travel  
506 along the rails faster. More specifically, they could submit one of three possible answers:  
507 ‘Wheel 1’, ‘Wheel 2’ or ‘No difference’. There were 10 items (i.e., 5 inertia items and 5  
508 centre of mass items; see Supplementary Methods). They could take as much time as  
509 they needed to complete the items. They had no feedback. All the participants saw the  
510 same items in the same order. The procedure for the analogous test and the transfer test  
511 was similar to Experiment 1, except that we modified the items so as to generate 12  
512 inertia items and 12 centre of mass items for each test (see Supplementary Methods). All  
513 the tests are available at <https://osf.io/athg5/>. Note that, for the three-option test of  
514 Experiment 2, the 10 items proposed (i.e., the 10 pairs of wheels) were not the same as  
515 those used in the original three-option test developed by Derex et al.<sup>16</sup>. As explained  
516 above, the wheel used in Experiment 2 behaved similarly to Derex et al.’s<sup>16</sup> wheel.  
517 However, as explained above, slight differences remained between the two physical  
518 systems. Thus, when we measured the different speeds of the wheels used by Derex et  
519 al.<sup>16</sup> in their three-option test with our own wheel system, the differences of speed (in

520 %) between pairs of wheels were not strictly equivalent to the differences of speed (in  
521 %) found by Derex et al.<sup>16</sup> with their wheel system. Therefore, we preferred to create 10  
522 new items (i.e., 10 new pairs of wheels), which were more equivalent in terms of speed  
523 differences (in %; see Supplementary Methods).

524 **Statistical analyses.** In both experiments, we first explored the wheel speed over  
525 generations in the experimental group. Wheel speed corresponded here to the best  
526 speed of the last two trials. Wheels that did not travel down were assigned a speed of 0  
527 m h<sup>-1</sup>. We used Bayesian applied regression modelling in R<sup>44</sup> (rstanarm package<sup>45</sup>) to fit  
528 a linear model with 'Wheel speed' as outcome variable, 'Generation' as fixed effect, and  
529 'Chain's identity' as random effect. Inferences were made using 95% credible intervals.  
530 The same analyses were conducted for the analogous and transfer scores in the  
531 experimental group (Experiments 1 and 2) as well as for the three-option score  
532 (Experiment 2). In addition, in both experiments, we used Bayesian applied regression  
533 modelling in R<sup>44</sup> (rstanarm package<sup>45</sup>) to explore links between variables of interest in  
534 the experimental group. Inferences were made using 95% credible intervals. We fitted a  
535 generalized linear mixed model with 'Variable of interest 1 (e.g., analogous score)' as  
536 outcome variable, 'Variable of interest 2 (e.g., wheel speed)' as fixed effect, and  
537 'Generation' and 'Chain's identity' as random effects. Correlational analyses and t-tests  
538 were carried out in R<sup>44</sup> with the bayestestR package<sup>46</sup>. In both experiments,  
539 correlational analyses were performed between the analogous and transfer scores for  
540 the control group. We also computed t-tests to compare the understanding scores  
541 (Analogous and Transfer, Experiments 1 and 2; Three-option, Experiment 2) between  
542 each generation in the experimental group and the control group.



543 **Code availability**

544 Codes used in this study are available at <https://osf.io/athg5/>.

545 **Data availability**

546 The data that support the findings of this study are available at <https://osf.io/athg5/>.

547 **References**

- 548 1. Boyd, R., & Richerson, P. J. Why culture is common but cultural evolution is rare. *Proc. Brit. Acad.* **88**,  
549 77–93 (1996).
- 550 2. Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: A  
551 comparative perspective. *Biol. Rev.* **89**, 234–301 (2014).
- 552 3. Henrich, J. *The secret of our success: How culture us driving human evolution, domesticating our*  
553 *species, and making use smarter* (Princeton Univ. Press, 2015).
- 554 4. Csibra, G., & Gergely, G. Natural pedagogy. *Trends Cogn. Sci.* **13**, 148–153 (2009).
- 555 5. Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. Social learning  
556 strategies: Bridge-building between fields. *Trends Cogn. Sci.* **22**, 651–665 (2018).
- 557 6. Tomasello, M., Carpenter, M. & Call, J. Understanding sharing intentions: The origins of cultural  
558 cognition. *Behav. Brain Sci.* **28**, 675–735 (2005).
- 559 7. Caldwell, C. A., & Millen, A. E. Social learning mechanisms and cumulative cultural evolution: Is  
560 imitation necessary? *Psychol. Sci.* **20**, 1478–1483 (2009).
- 561 8. Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B. & Laland, K. N. Identification of the social and  
562 cognitive processes underlying human cumulative culture. *Science* **335**, 1114–1118 (2012).
- 563 9. Caldwell, C. A., Renner, E., & Atkinson, M. Human teaching and cumulative cultural evolution. *Rev.*  
564 *Phil. Psychol.* **9**, 751–770 (2018).
- 565 10. Herrmann, E., Call, J., Hernandez-Lloreda, M., Hare, B., & Tomasello, M. Humans have evolved  
566 specialized skills of social cognition: The cultural intelligence hypothesis. *Science* **317**, 1360–1366  
567 (2007).
- 568 11. Heyes, C. Who knows? Metacognitive social learning strategies. *Trends Cogn. Sci.* **20**, 204–213  
569 (2016).
- 570 12. Whiten, A. Social learning and culture in child and chimpanzee. *Annu. Rev. Psychol.*, **68**, 129–154  
571 (2017).
- 572 13. Tomasello, M. *Becoming human. A theory of ontogeny* (Belknap Press, 2019).
- 573 14. Boyd, R., Richerson, P. J. & Henrich, J. The cultural niche: Why social learning is essential for human  
574 adaptation. *Proc. Natl Acad. Sci. USA* **108**, 10918–10925 (2011).
- 575 15. Derex, M., & Boyd, R. Technical reasoning alone does not take humans this far. *Behav. Brain Sci.* **43**,  
576 e156 (2020).
- 577 16. Derex, M., Bonnefon, J. F., Boyd, R., & Mesoudi, A. Causal understanding is not necessary for the  
578 improvement of culturally evolving technology. *Nature Hum. Behav.* **3**, 446–452 (2019)
- 579 17. Osiurak, F., & Reynaud, E. The elephant in the room: What matters cognitively in cumulative  
580 technological culture. *Behav. Brain Sci.* **43**, e156 (2020).
- 581 18. Reynaud, E., Lesourd, M., Navarro, J., & Osiurak, F. On the neurocognitive origins of human tool use.  
582 A critical review of neuroimaging data. *Neurosci. BioBehav. Rev.* **64**, 421–437 (2016).
- 583 19. Reynaud, E., Navarro, J., Lesourd, M., & Osiurak, F. To watch is to work: A critical review of  
584 neuroimaging data on Tool-use Observation Network (ToON). *Neuropsychol. Rev.* **29**, 484–497  
585 (2019).

- 586 20. Osiurak, F., De Oliveira, E., Navarro, J., Lesourd, M., Claidière, N., & Reynaud, E. Physical intelligence  
587 does matter to cumulative technological culture. *J. Exp. Psychol. Gen.* **145**, 941–948 (2016).
- 588 21. De Oliveira, E., Reynaud, E., & Osiurak, F. Roles of technical reasoning, theory of mind, creativity,  
589 and fluid cognition in cumulative technological culture. *Hum. Nat.* **30**, 326–340 (2019).
- 590 22. Osiurak, F., De Oliveira, E., Navarro, J., & Reynaud, E. The castaway island: Distinct roles of theory of  
591 mind and technical reasoning in cumulative technological culture. *J. Exp. Psychol. Gen.* **149**, 58–66  
592 (2020).
- 593 23. Derex, M., & Boyd, R. The foundations of the human cultural niche. *Nature Comm.* **6**, 8398 (2015).
- 594 24. Derex, M., & Boyd, R. Social information can potentiate understanding despite inhibiting cognitive  
595 effort. *Sci. Rep.* **8**, 9980 (2018).
- 596 25. Whiten, A., Horner, V. & de Waal, F. B. M. Conformity to cultural norms of tool use in chimpanzees.  
597 *Nature* **437**, 737–740 (2005).
- 598 26. Claidière, N., & Whiten, A. Integrating the study of conformity and culture in humans and  
599 nonhuman animals. *Psychol. Bull.* **138**, 126–145 (2012).
- 600 27. Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G.,  
601 Wrangham, R. W., & Boesch, C. Cultures in chimpanzees. *Nature* **399**, 682–685 (1999).
- 602 28. Horner, V., Whiten, A., Flynn, E., & de Waal, F. B. M. Faithful replication of foraging techniques along  
603 cultural transmission chains by chimpanzees and children. *Proc. Natl Acad. Sci. USA* **103**, 13878–  
604 13883 (2006).
- 605 29. Gruber, T., Muller, M. N., Strimling, P., Wrangham, R. W., & Zuberbühler, K. Wild chimpanzees rely  
606 on cultural knowledge to solve an experimental honey acquisition task. *Curr. Biol.* **19**, 1806–1810  
607 (2009).
- 608 30. Gruber, T. Great apes do not learn novel tool use easily: Conservatism, functional fixedness, or  
609 cultural influence? *Int. J. Primatol.* **37**, 296–316 (2016).
- 610 31. Tennie, C., Call, J., & Tomasello, M. Push or pull: Imitation vs. emulation in great apes and human  
611 children. *Ethology* **112**, 1159–1169 (2006).
- 612 32. Bandini, E., & Tennie, C. Spontaneous reoccurrence of “scooping”, a wild tool-use behaviour, in  
613 naïve chimpanzees. *PeerJ* **5**, e3814 (2017).
- 614 33. Clay, Z., & Tennie, C. Is overimitation a uniquely human phenomenon? Insights from human  
615 children as compared to bonobos. *Child Development* **89**, 1535–1544 (2018).
- 616 34. Whiten, A., Horner, V. & Marshall-Pescini, S. R. J. Cultural panthropology. *Evol. Anthropol.* **12**, 92–  
617 105 (2003).
- 618 35. Martin-Ordas, G., Call, J., & Colmenares, F. Tubes, tables and traps: Great apes solve two functionally  
619 equivalent trap tasks but show no evidence of transfer across tasks. *Anim. Cogn.* **11**, 423–430  
620 (2008).
- 621 36. Logan, C. J., Breen, A. J., Taylor, A. H., Gray, R. D., & Hoppitt, J. E. How New Caledonian crows solve  
622 novel foraging problems and what it means for cumulative culture. *Learn. Behav.* **44**, 18–28 (2016).
- 623 37. Gruber, R., Schiestl, M., Boeckle, M., Frohnwieser, A., Miller, R., Gray, R. D., Clayton, N. S. & Taylor, A.  
624 H. New Caledonian crows use mental representations to solve metatool problems. *Curr. Biol.* **29**,  
625 686–692 (2019).
- 626 38. Jelbert, S. A., Hosking, R. J., Taylor, A. H., & Gray, R. D. Mental template matching is a potential  
627 cultural transmission mechanism for New Caledonian crow tool manufacturing traditions. *Sci. Rep.*  
628 **8**, 8956 (2018).
- 629 39. Moll, H., & Tomasello, M. Cooperation and human cognition: The Vygotskian intelligence  
630 hypothesis. *Phil. Trans. R. Soc. B* **362**, 639–648 (2007).
- 631 40. Baillargeon, R., Needham, A., & DeVos, J. The development of young infants’ intuitions about  
632 support. *Early Dev. Parent.* **1**, 69–78 (1992).
- 633 41. Pinker, S. The cognitive niche: Coevolution of intelligence, sociality, and language. *Proc. Natl Acad.*  
634 *Sci. USA* **107**, 8993–8999 (2010).
- 635 42. Vaesen, K. The cognitive bases of human tool use. *Behav. Brain Sci.* **35**, 203–218 (2012).
- 636 43. Heyes, C. Enquire within: Cultural evolution and cognitive science. *Phil. Trans. R. Soc. B* **373**,  
637 20170051 (2018).

- 638 44. R Development Core Team *R: A language and environment for statistical computing* (R Foundation  
639 for Statistical Computing, 2011).  
640 45. Gabry, J., & Goodrich, B. rstanarm: Bayesian applied regression modeling via Stan. R package  
641 version 2.15.3 (Retrieved from <http://mc-stan.org/rstanarm/>, 2017).  
642 46. Makowski, D., Ben-Shachar, M. S., & Lüdtke, D. bayestestR: Describing effects and their  
643 uncertainty, existence and significance within the Bayesian framework. *J. Open Source Softw.* **4**,  
644 1541.

## 645 **Acknowledgments**

646 We thank Tim Pownall for the English proofreading of this manuscript, and Vivien  
647 Gaujoux, Gabriel Jarjat and Nicolas Baltenneck for assistance in data collection. This  
648 work was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042)  
649 of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11- IDEX-  
650 0007) operated by the French National Research Agency (ANR).

## 651 **Author contributions**

652 F.O. developed the research question. F.O., S.L., J.A., J.B., A.B., J.N., and E.R. helped design  
653 the experimental task and protocol. J.B. built the wheel system. S.L., J.A. and F.O.  
654 performed the experiments. F.O. and E.R. analysed the data. F.O. wrote the manuscript  
655 with input from S.L., J.A., J.B., A.B., J.N., and E.R.

## 656 **Competing interests**

657 The authors declare that there are no financial or non-financial competing interests.

## 658 **Additional information**

659 **Supplementary information** is available for this paper at XX.

660 **Correspondence and requests for materials** should be addressed to F.O.

661 **Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

662

663 **Figure Legends**

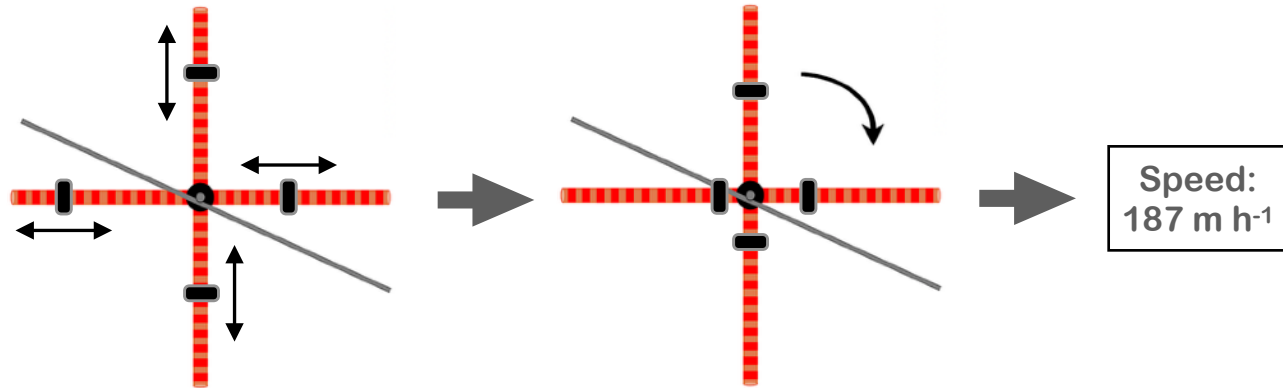
664 **Fig. 1 | Experimental task and design.** **a.** Illustration of the wheel system used in Experiments 1 and 2. For each trial,  
665 the participants could move the four weights to optimize the speed of the wheel. **b.** In both experiments, each  
666 participant in the experimental group performed the task as a member of a chain of 5 participants (i.e., 5 generations).  
667 They had five trials to optimize the speed of the wheel. Each participant (except those of the last generation)  
668 transmitted their last two weight configurations and the associated speeds (grey) to the next participant. After the  
669 five trials, the participants in the experimental group completed the understanding tests consisting in choosing the  
670 wheel that would travel down faster. The participants in the control group (i.e., no experience with the wheel system)  
671 also completed the understanding tests. There were two understanding tests (Analogous and Transfer; **c** and **d**) in  
672 Experiment 1 and three (Three-option, Analogous, and Transfer) in Experiment 2. The design of the analogous and  
673 transfer tests differed between Experiment 1 and Experiment 2 (see below).

674 **Fig. 2 | Parallel improvement of the wheel system and of its understanding (Experiment 1).** **a.** Wheel speed over  
675 generations for non-failure wheels (blue), and number of failures (wheels that did not descend; red). **b.**  
676 Understanding scores (Analogous: yellow; Transfer: green) over generations. The scores obtained on both tests by the  
677 control group are also shown. Links between the wheel speed, and the analogous score (**c**) and the transfer score (**d**)  
678 as well as between the two understanding scores in the experimental group (**e**) and the control group (**f**).

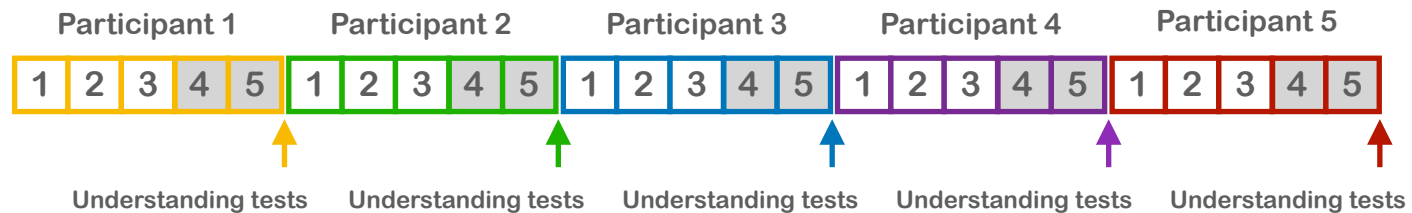
679 **Fig. 3 | Parallel improvement of the wheel system and of its understanding (Experiment 2).** **a.** Wheel speed over  
680 generations for non-failure wheels (blue), and number of failures (wheels that did not descend; red). **b.**  
681 Understanding scores (Analogous: yellow; Transfer: green) over generations. **c.** Understanding scores (Three-option  
682 test) over generations. The scores obtained on the three tests by the control group are also shown. Links between the  
683 wheel speed, and the analogous score (**d**) and the transfer score (**e**) as well as between these two understanding  
684 scores in the experimental group (**f**) and the control group (**g**). The link between the wheel speed and the three-option  
685 score is also shown in **h**.

686 **Fig. 4 | Links between the centre of mass and inertia scores in the three understanding tests for the**  
687 **participants in the experimental group (Experiment 2).** **a.** The centre of mass and inertia scores were negatively  
688 linked for the three-option test (Centre of mass score 95% credible interval: -0.80 to -0.14, median = -0.47; BF = 1.67).  
689 **b.** and **c.** A positive relationship was also found between the number of 'no difference' answers on centre of mass  
690 items and the inertia score (Inertia score 95% credible interval: 0.12 to 0.43, median = 0.27, BF = 16.00) and between  
691 the number of 'no difference' answers on inertia items and the centre of mass score (Centre of mass score 95%  
692 credible interval: 0.08 to 0.43, median = 0.25, BF = 1.78). **d.** and **e.** No statistical relationship was reported between  
693 the centre of mass and inertia scores for the analogous test (Centre of mass score 95% credible interval: -0.54 to 0.28,  
694 median = -0.12, BF = 0.06) and for the transfer test (Centre of mass score 95% credible interval: -0.40 to 0.31, median  
695 = -0.04, BF = 0.05).

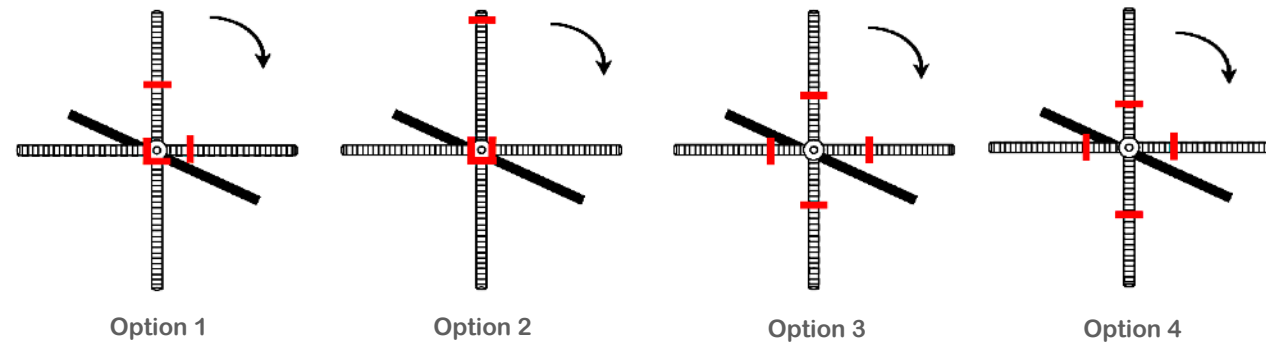
a. Trial



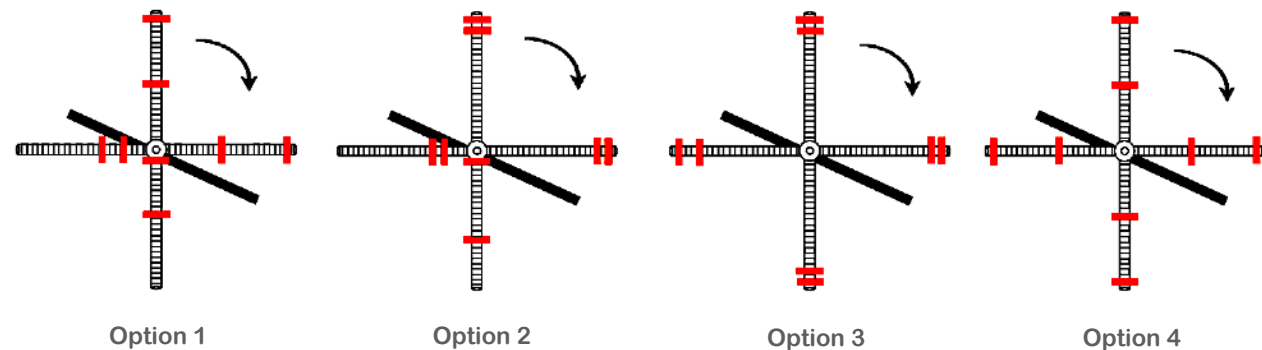
b. Chain

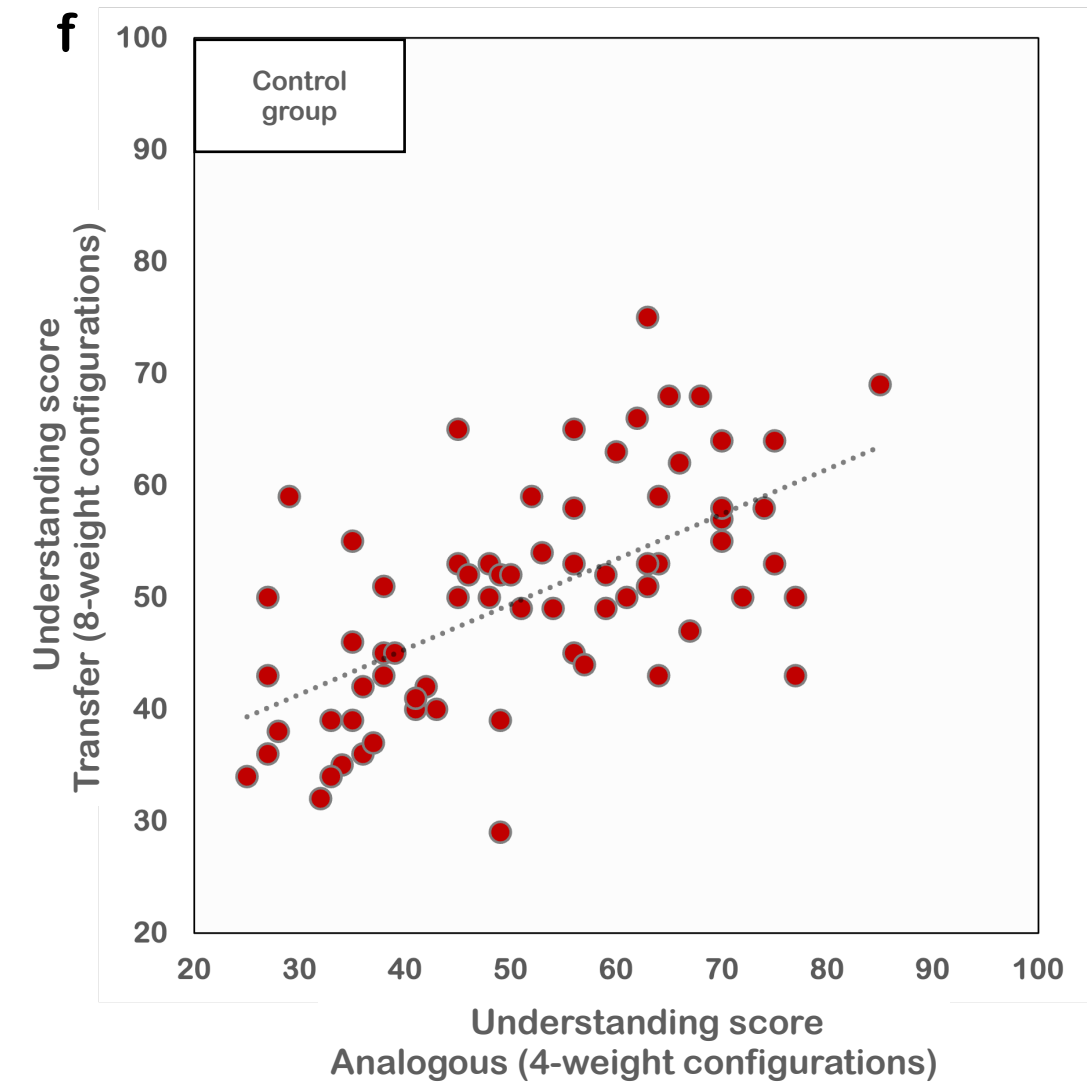
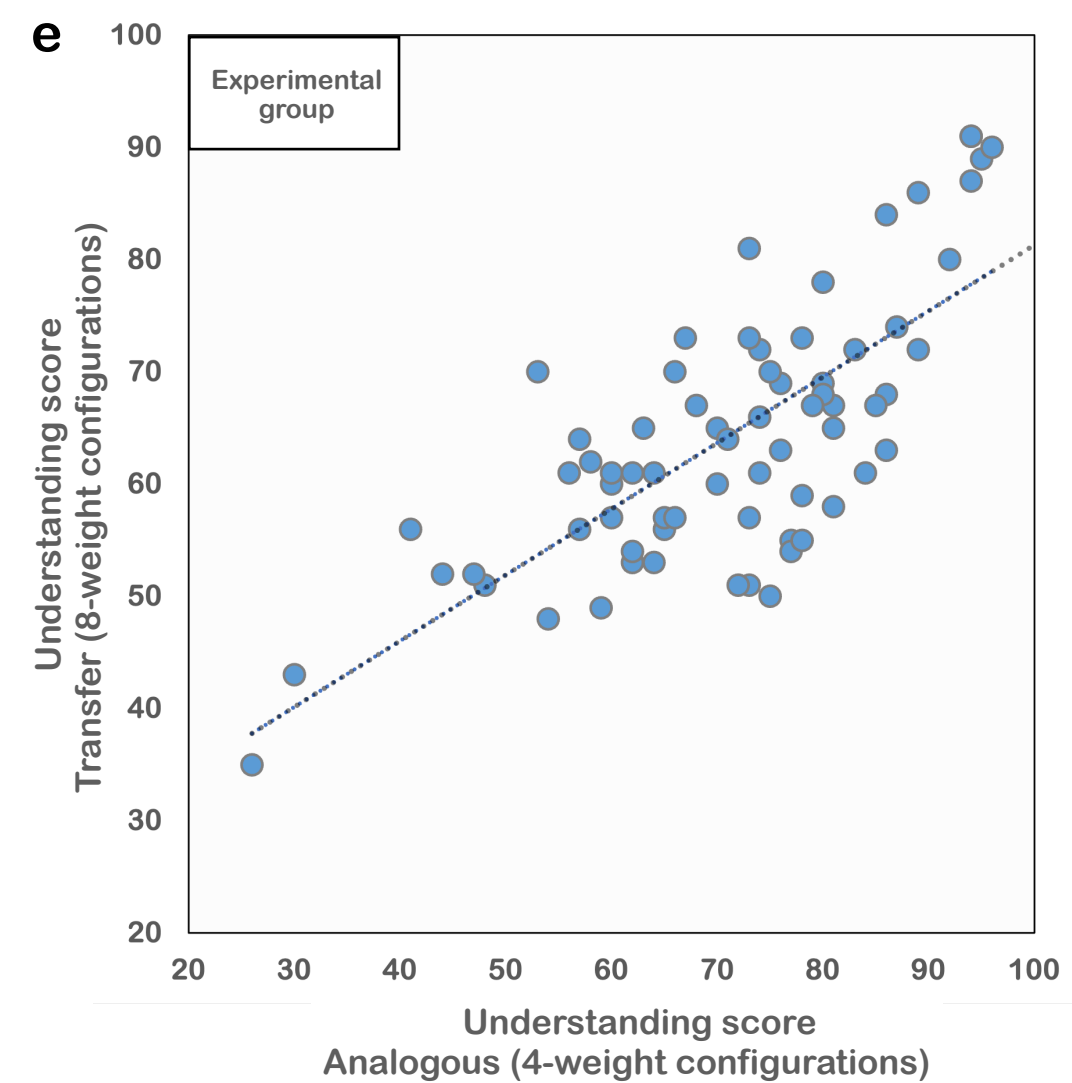
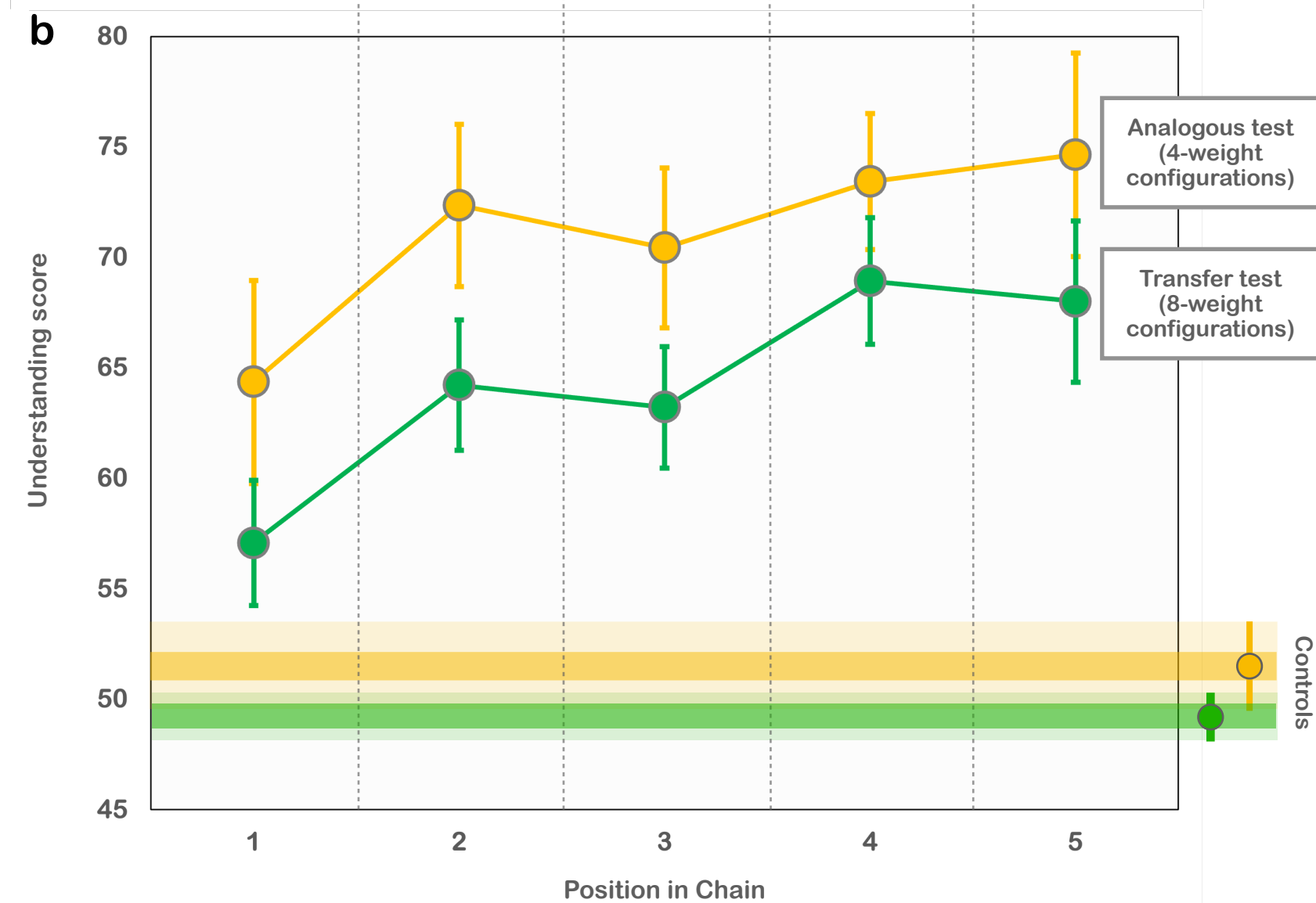
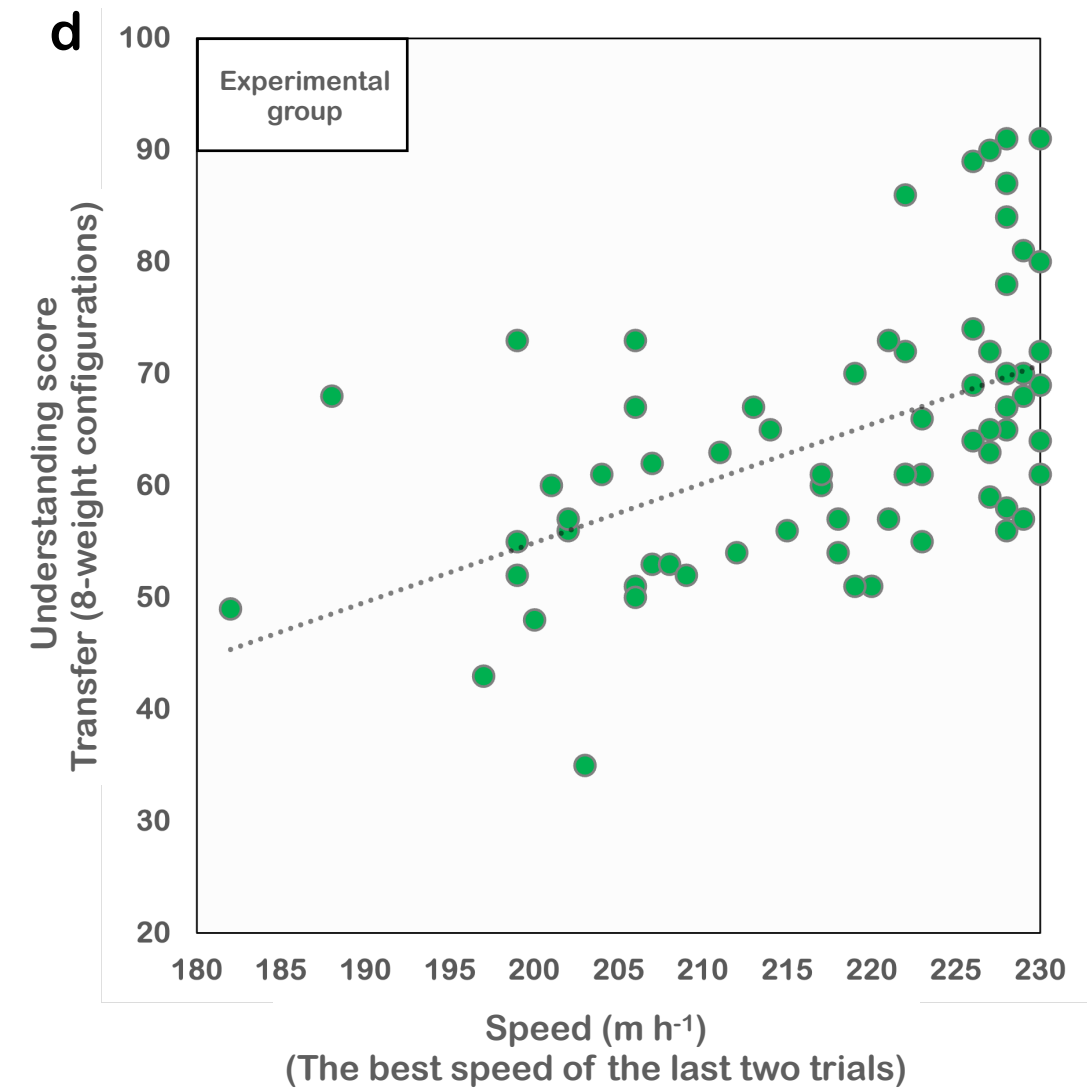
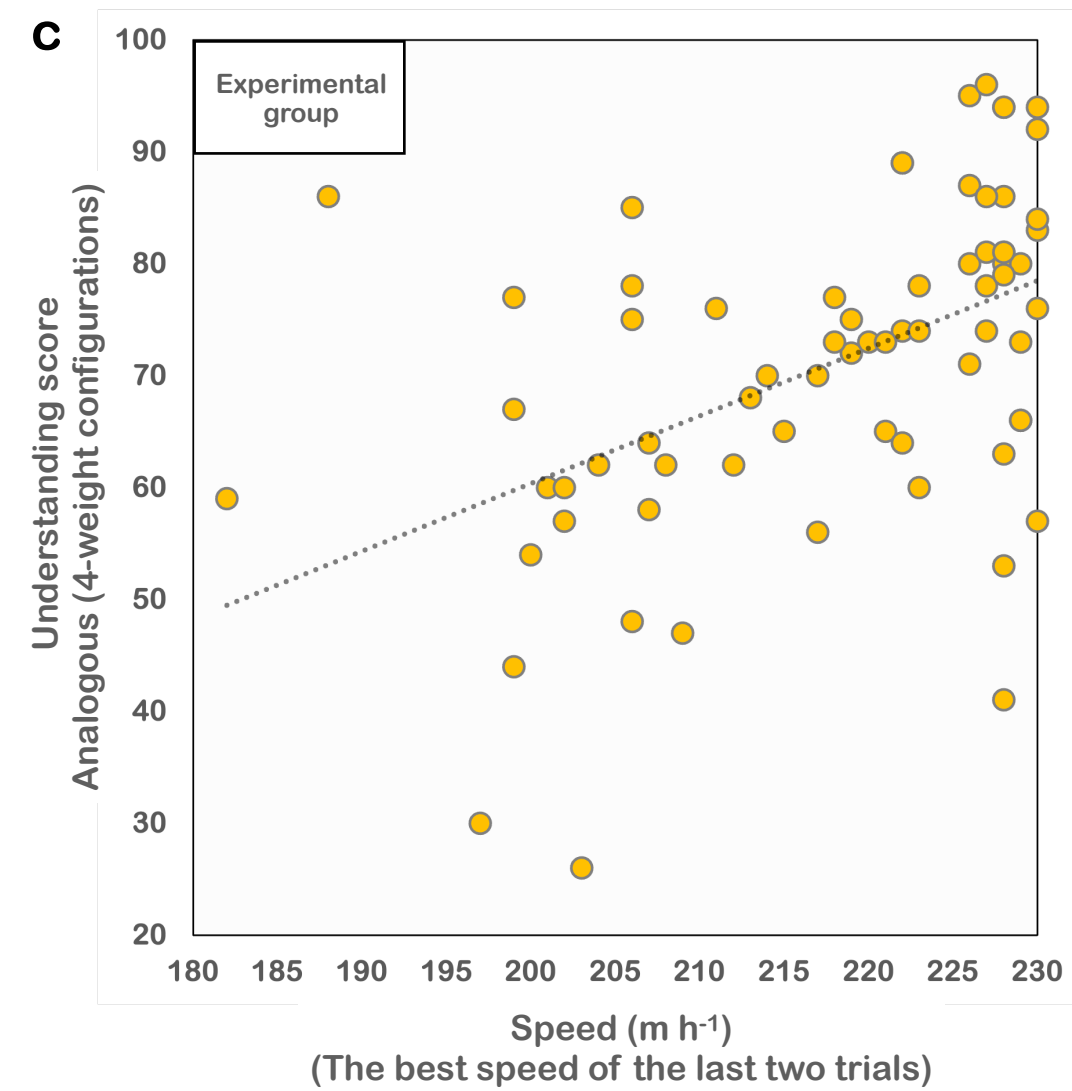
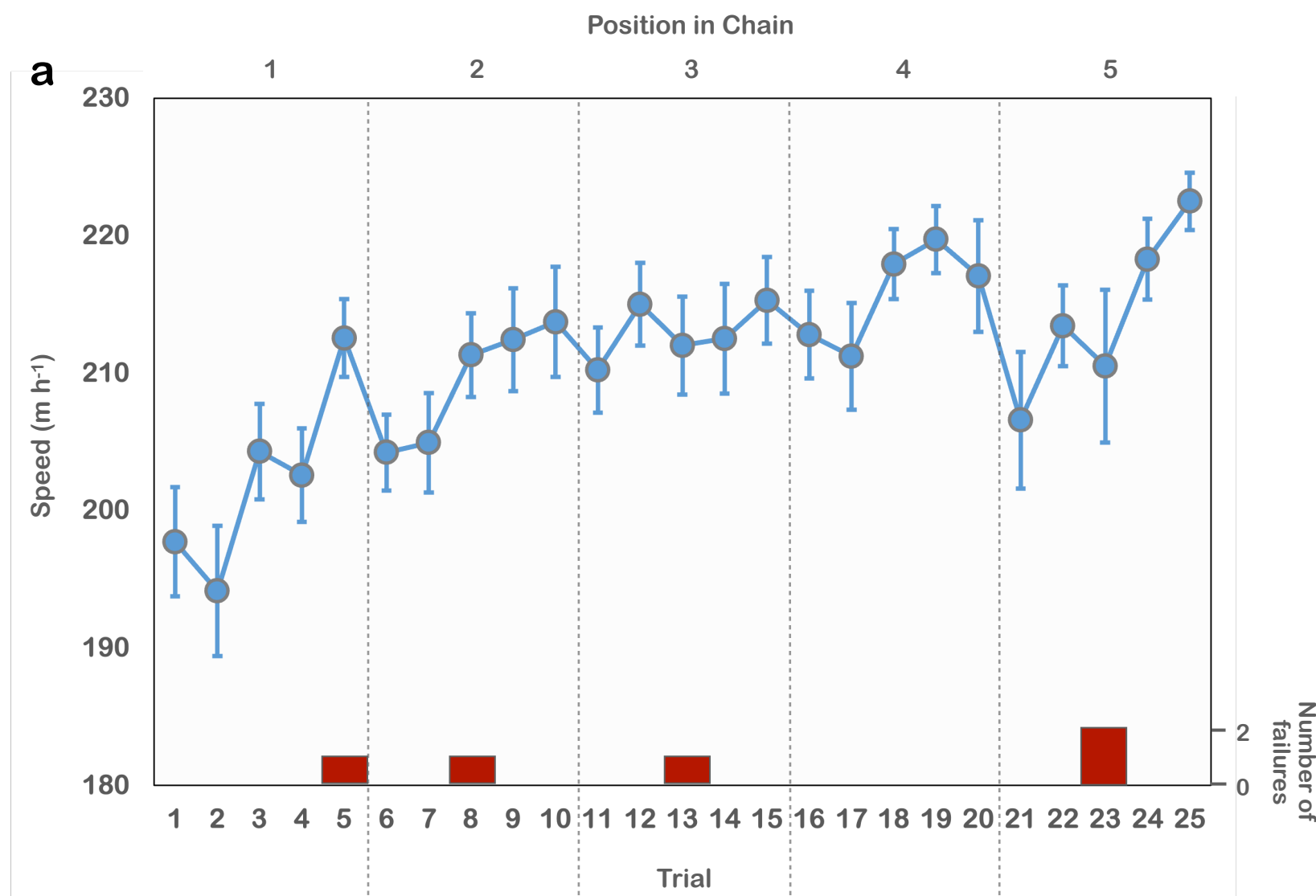


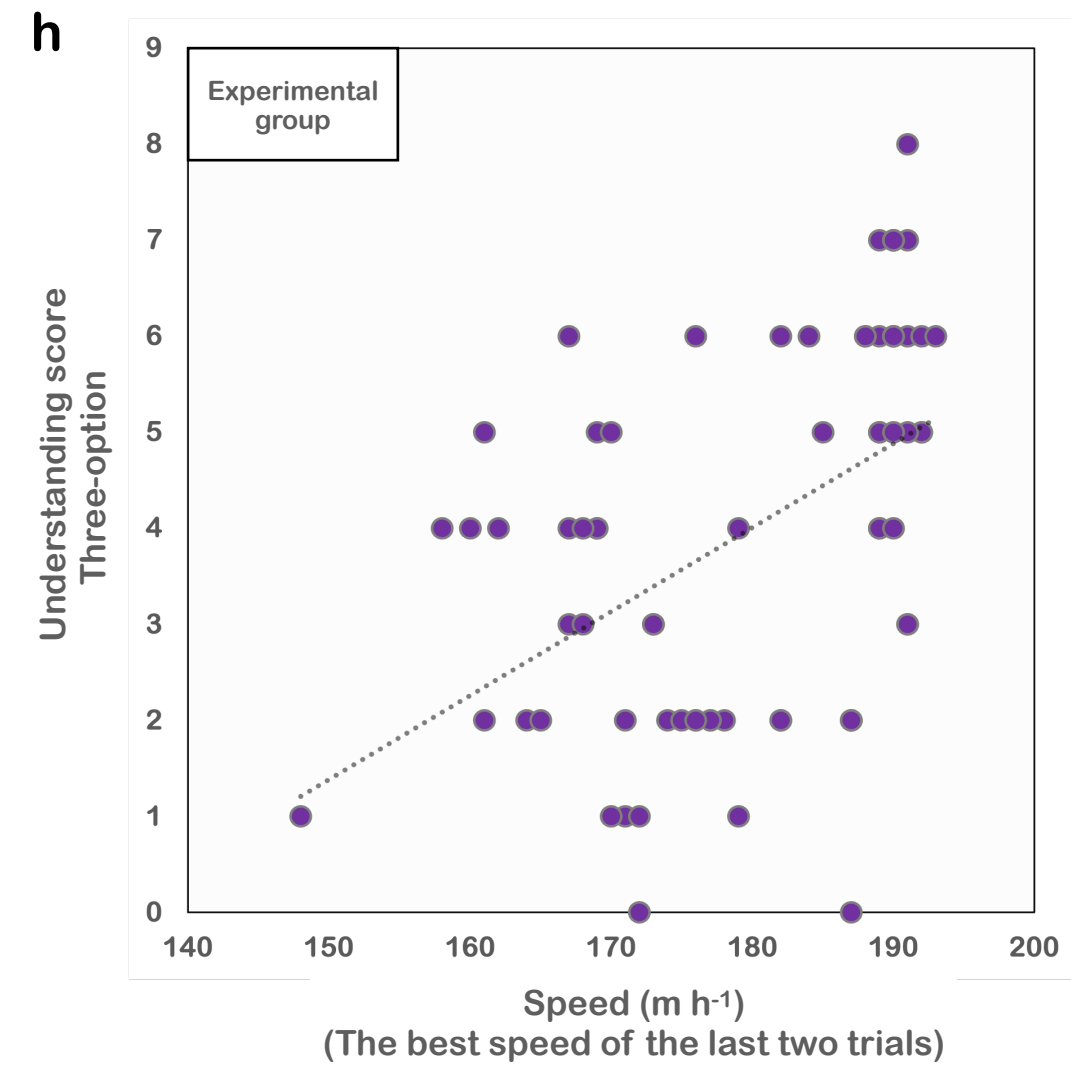
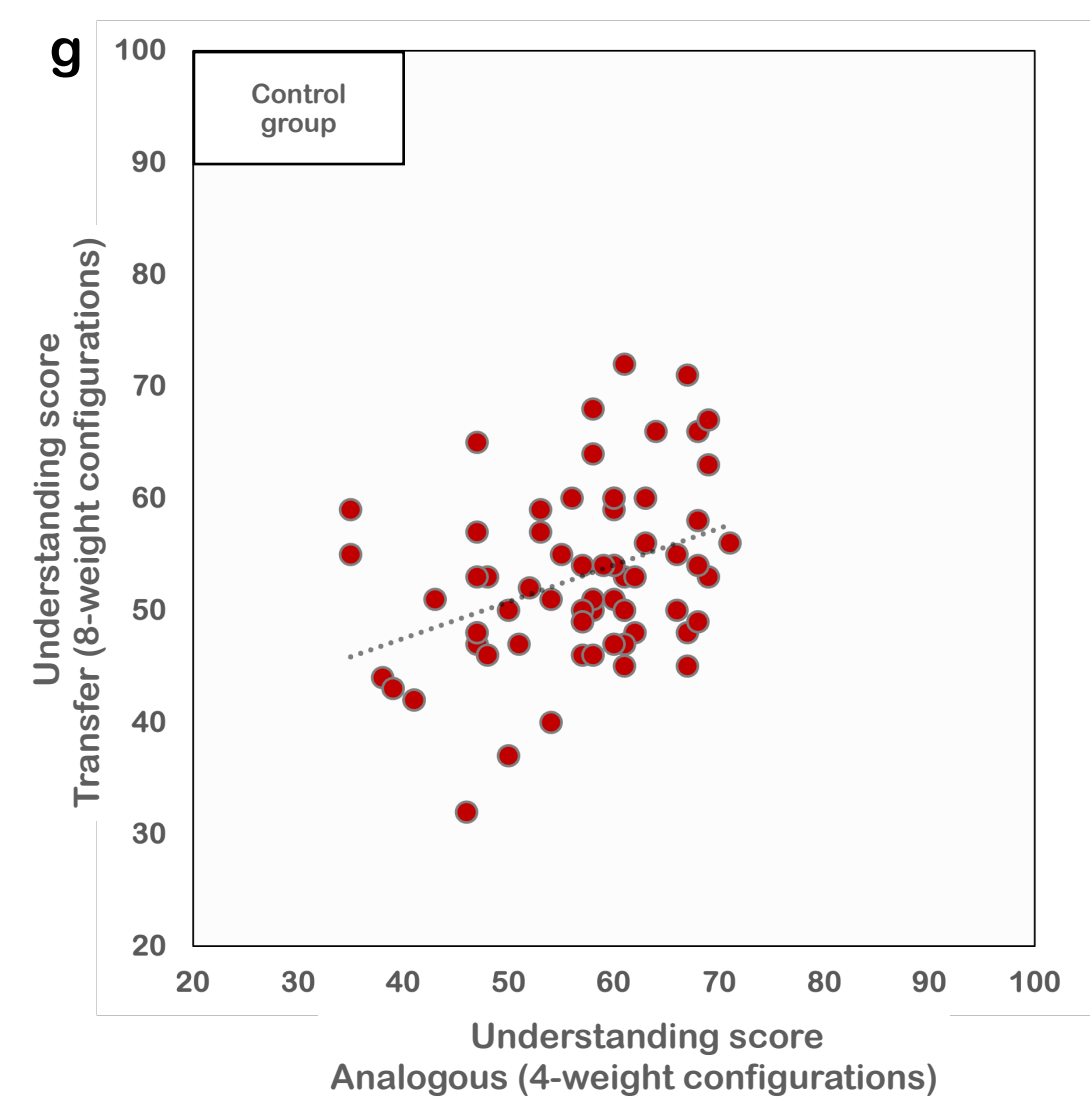
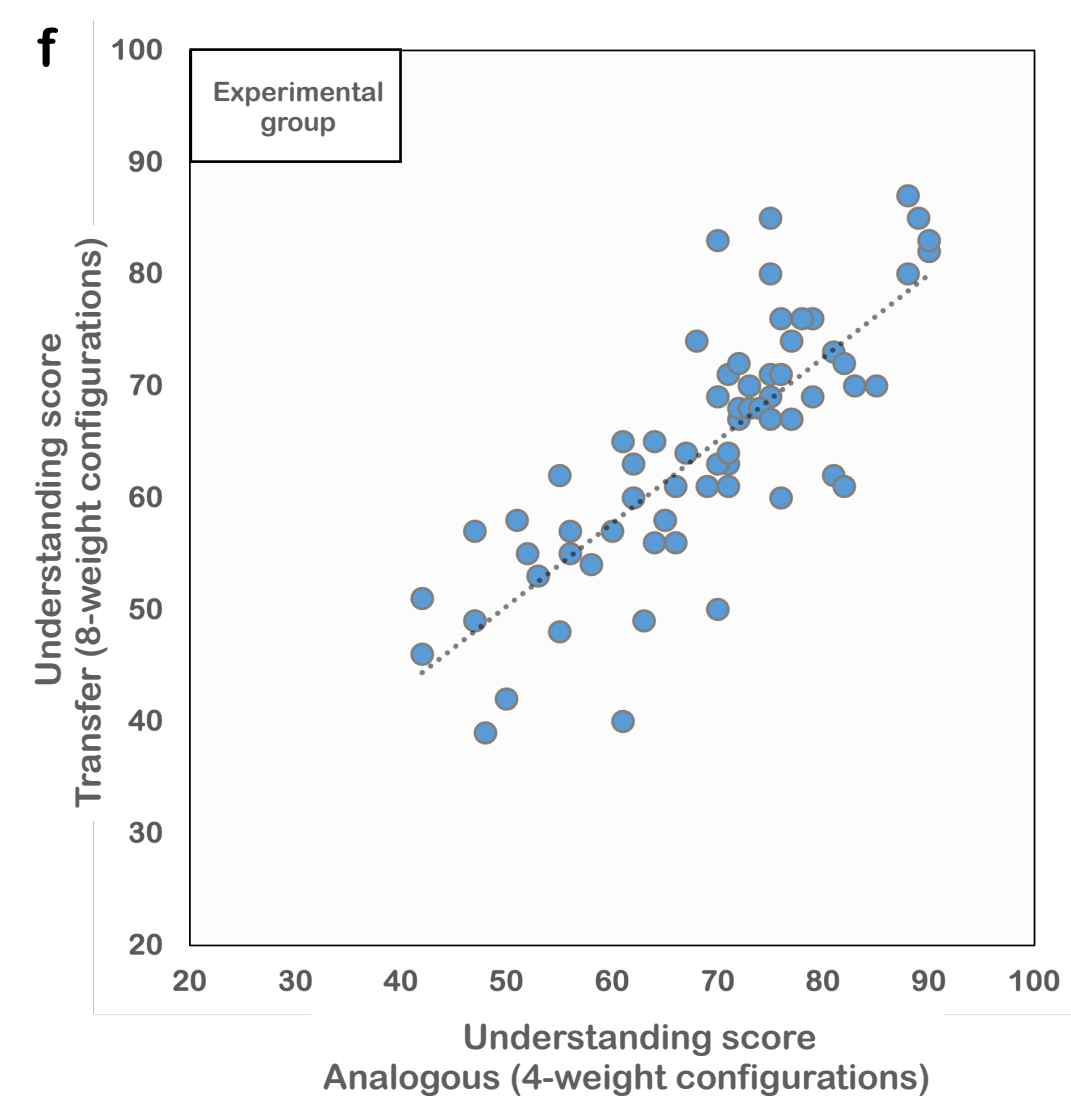
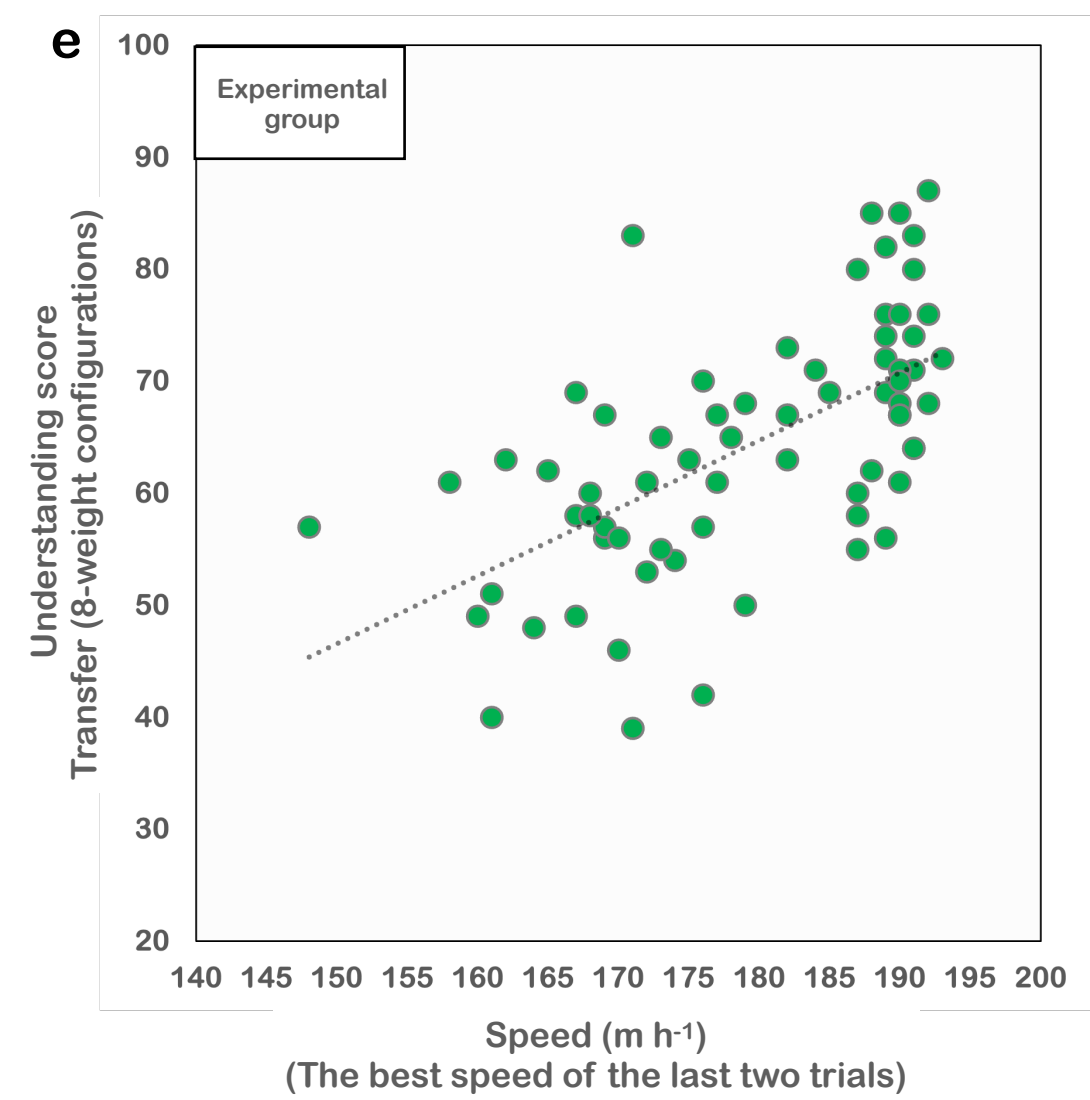
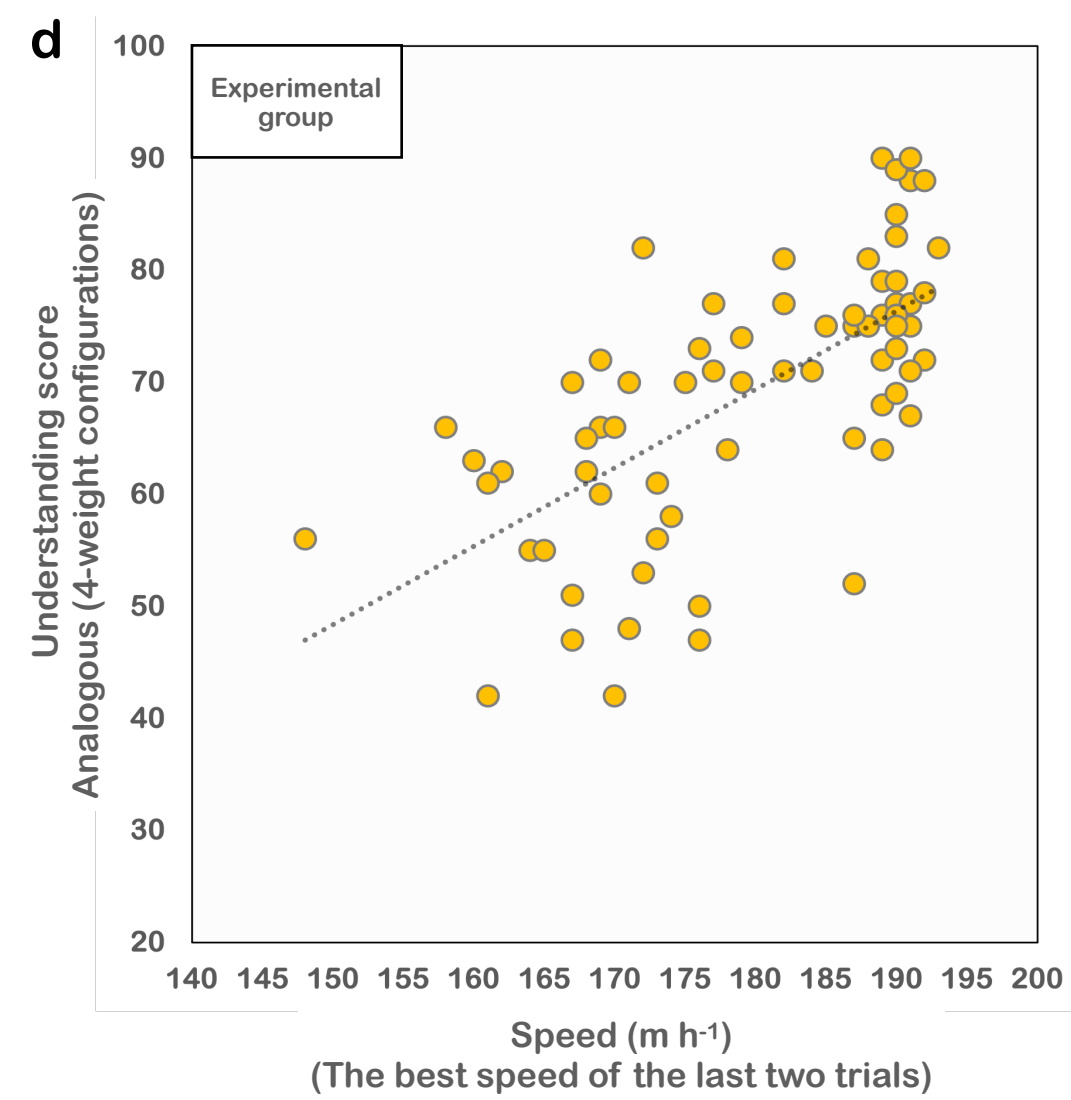
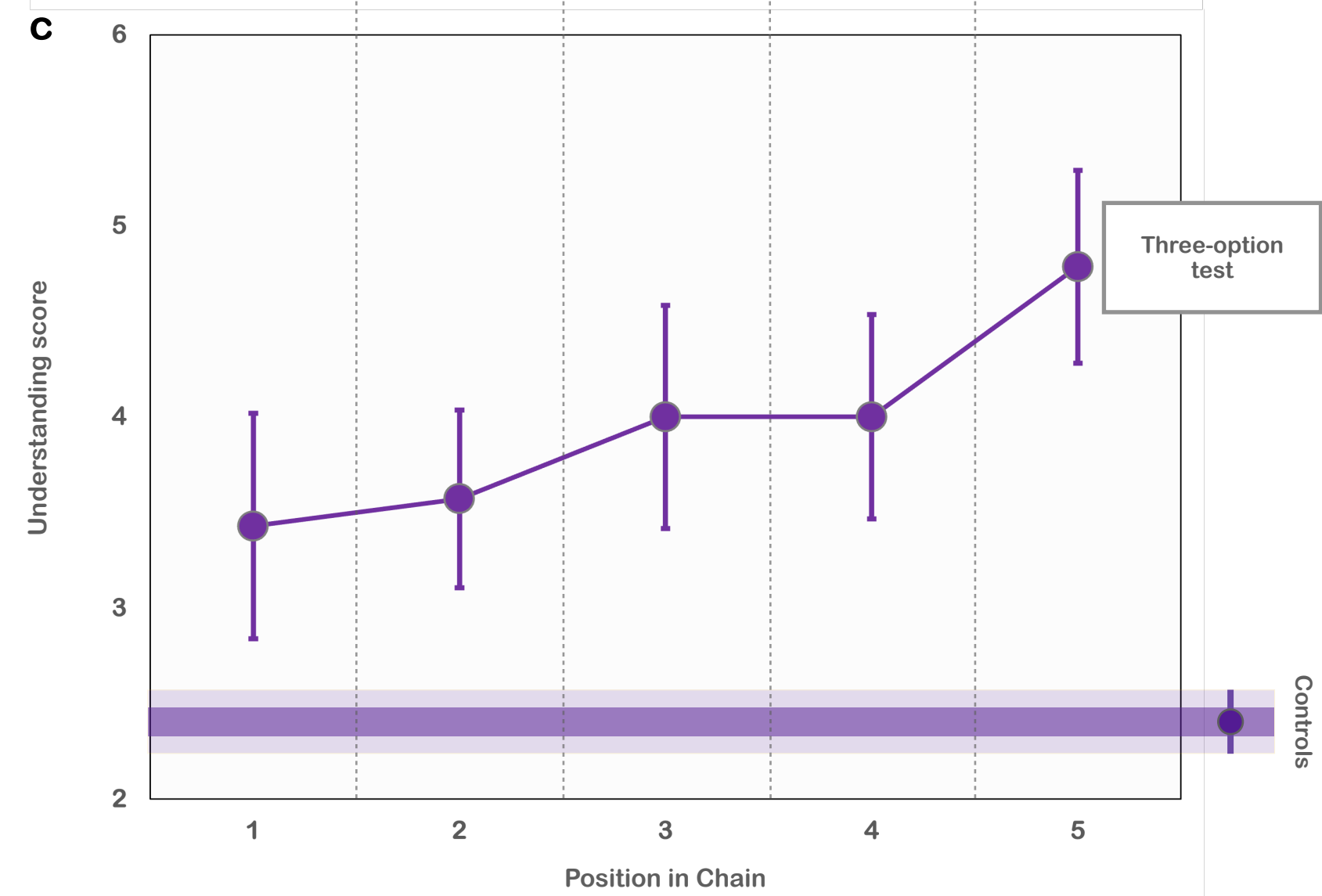
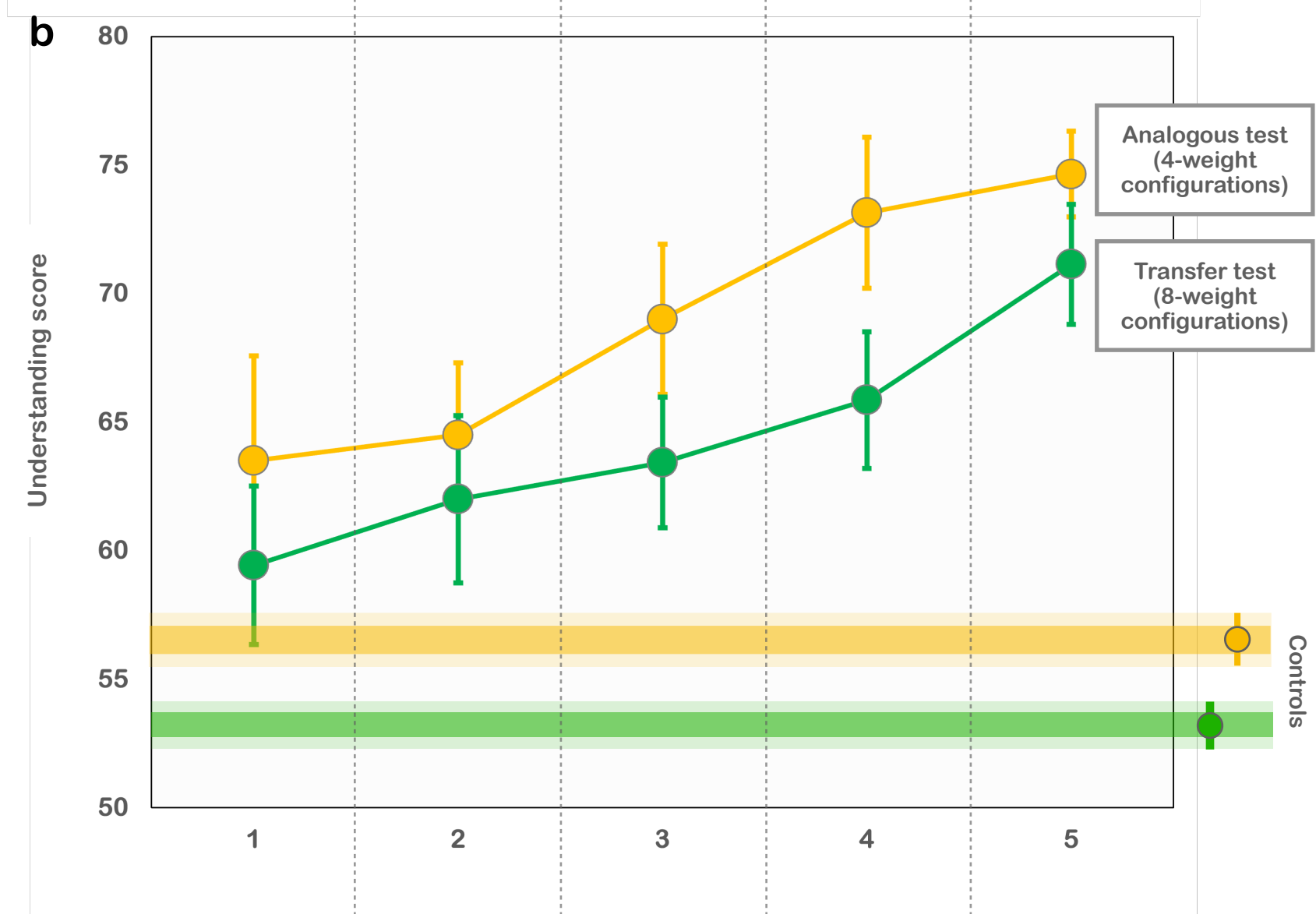
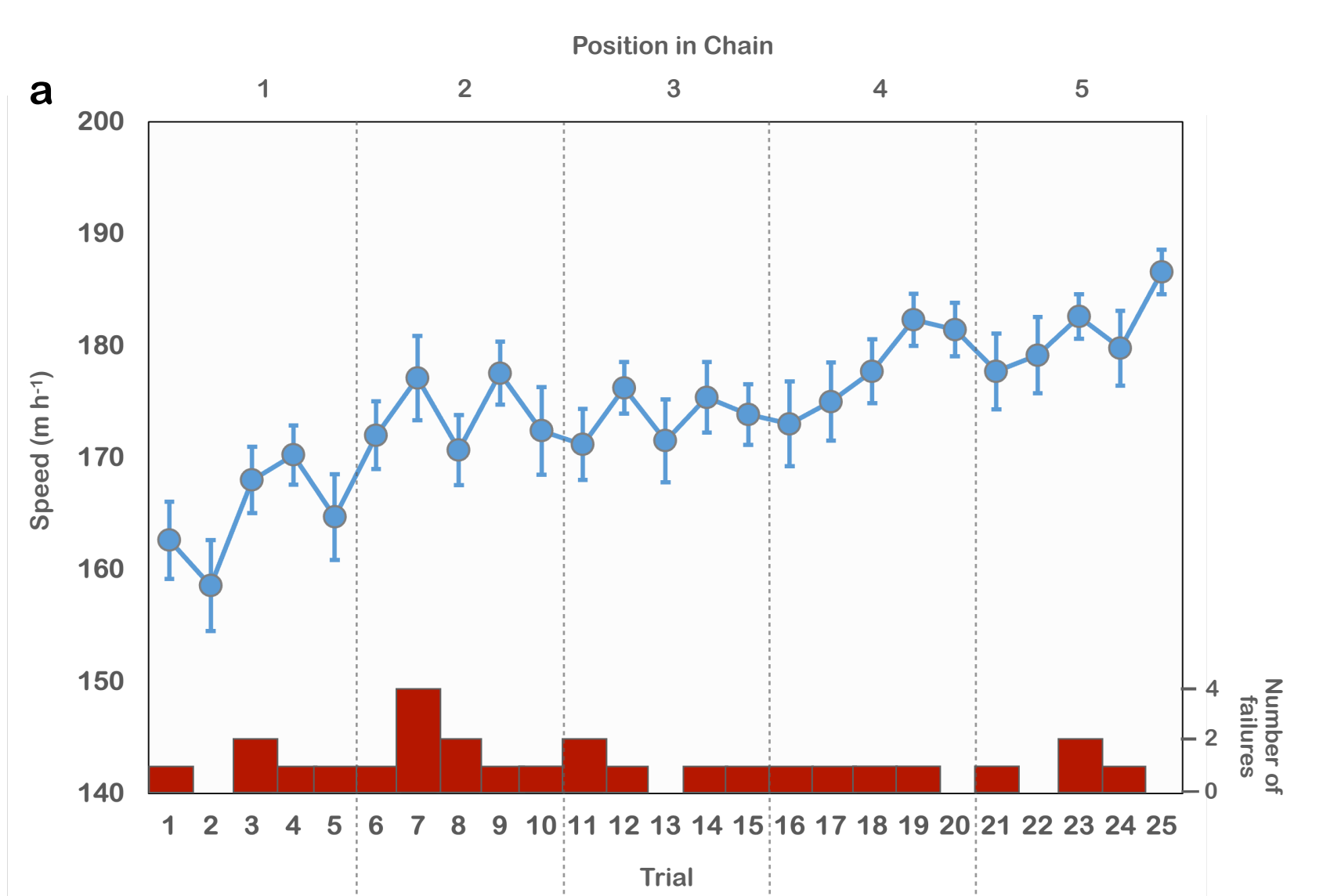
c. Understanding test: Analogous (4-weight configurations)



d. Understanding test: Transfer (8-weight configurations)

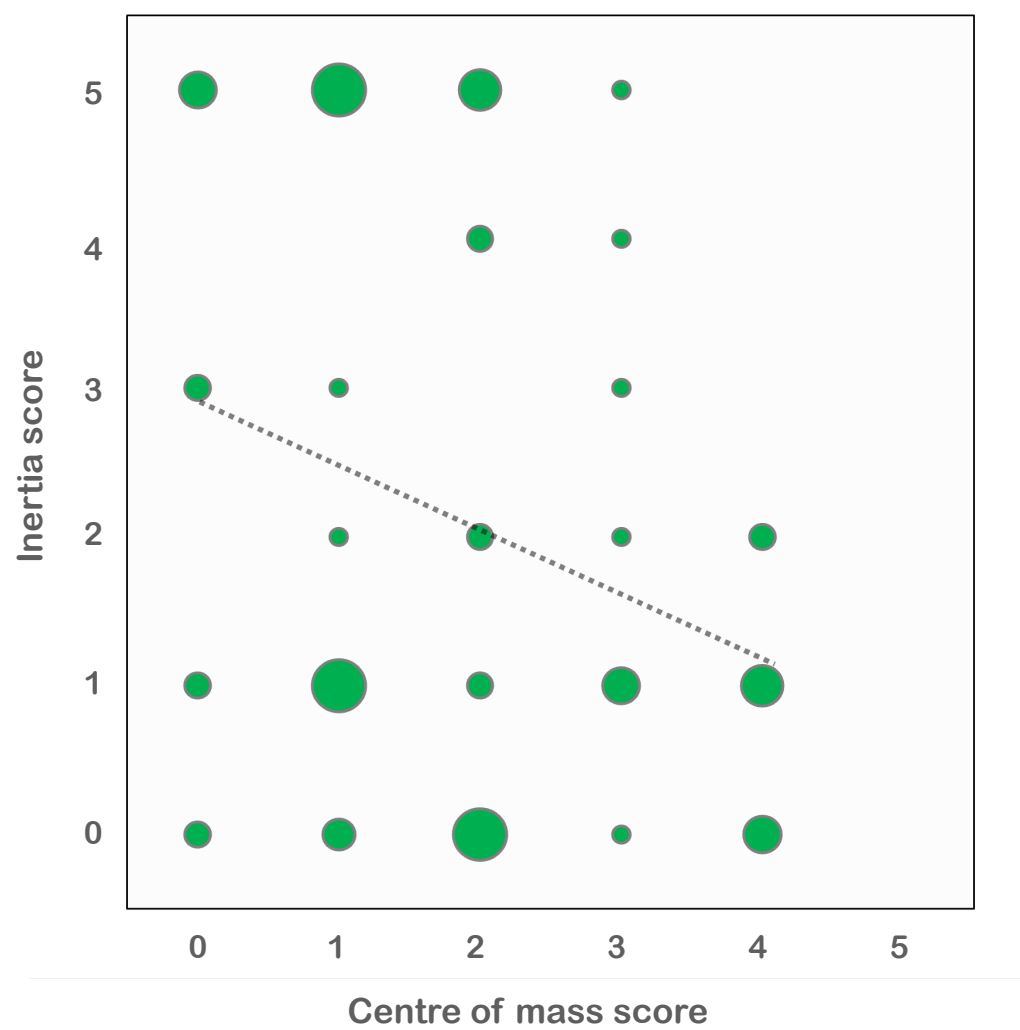






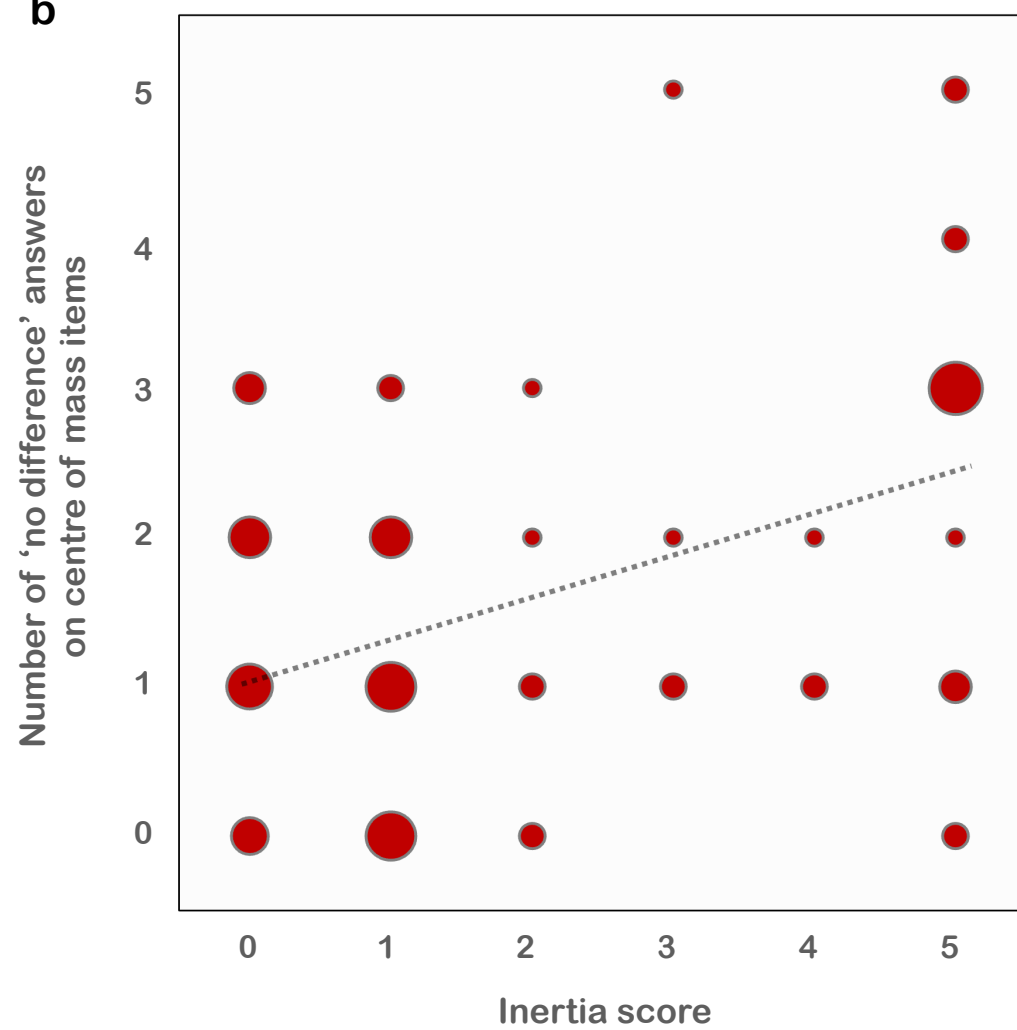
Three-option test

a



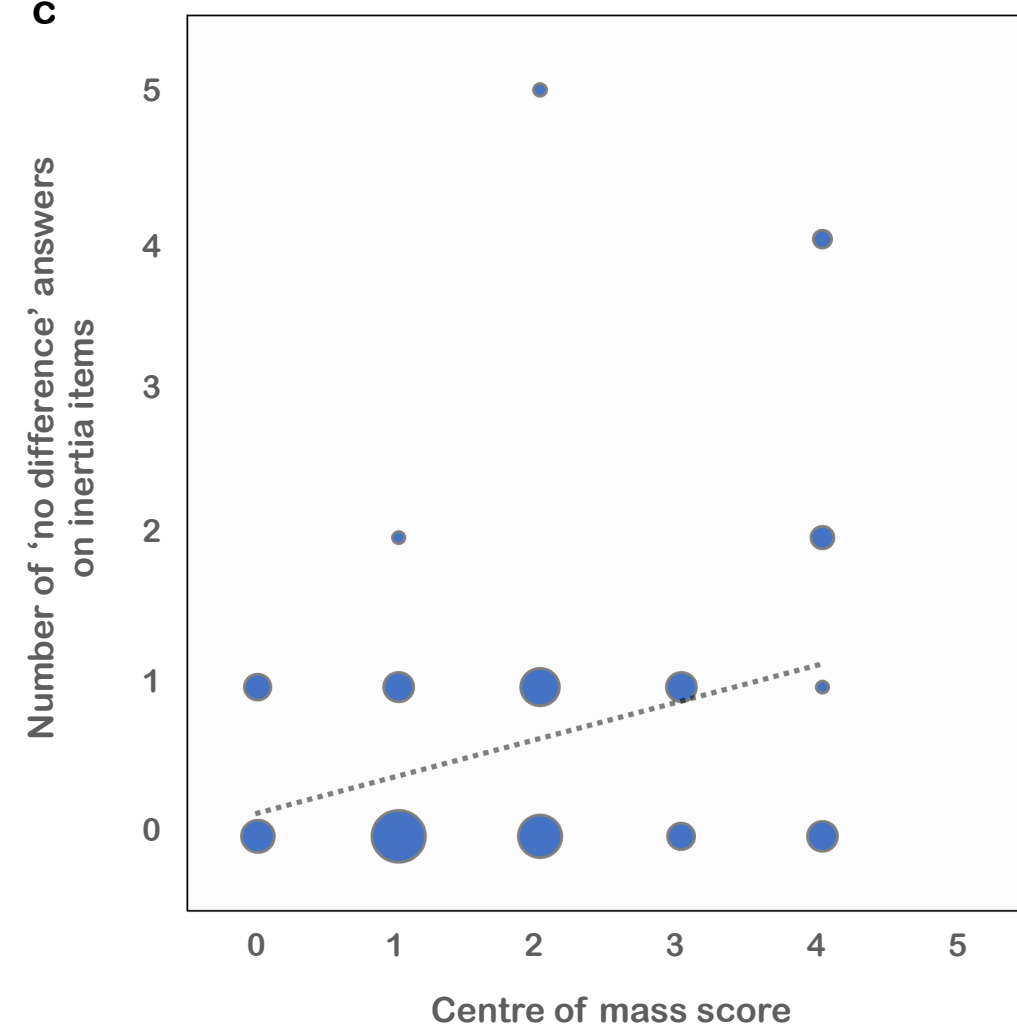
Three-option test

b



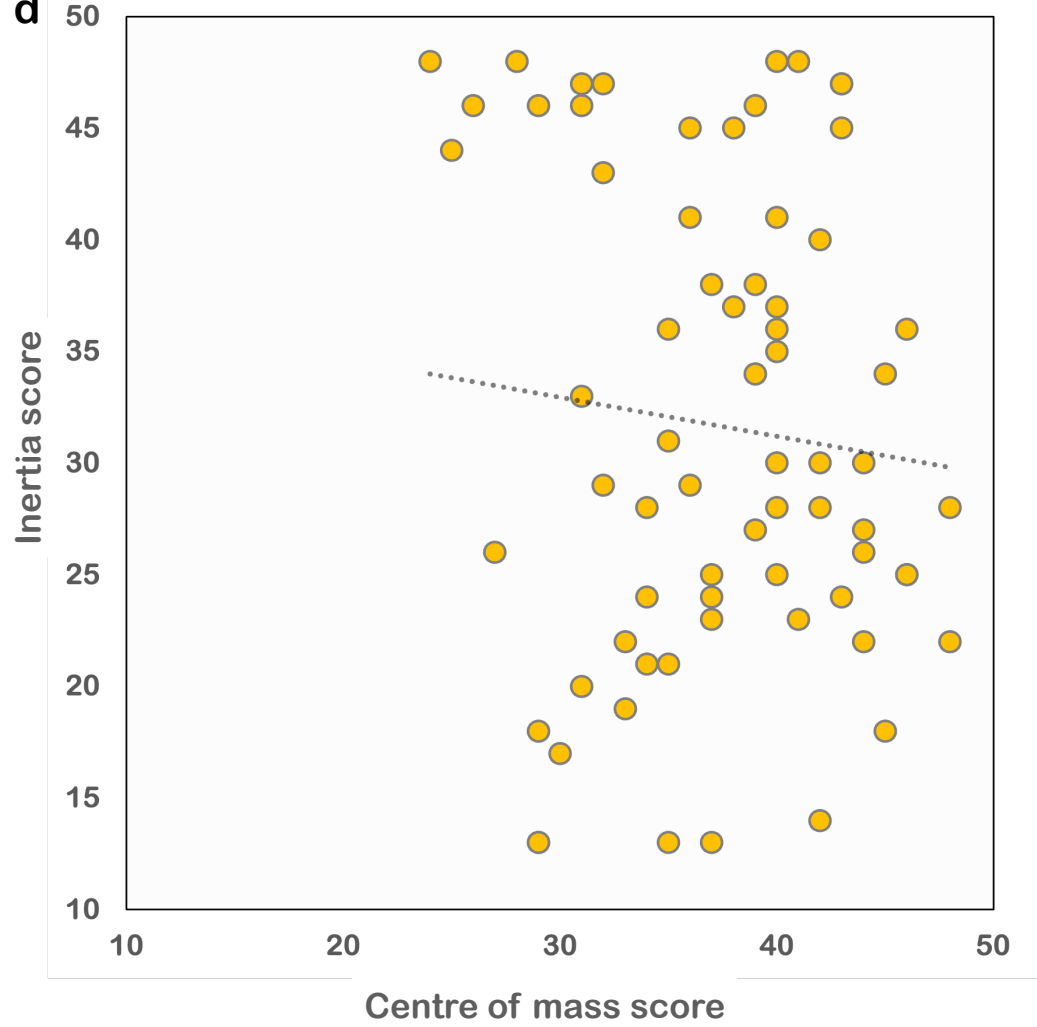
Three-option test

c



Analogous test

d



Transfer test

e

