

# The emergence and cultural persistence of ineffective solutions

Ali Seyhun Saral<sup>1,\*</sup>, Manvir Singh<sup>2,\*</sup>, Antoine Jacquet<sup>3</sup>, Haneul Jang<sup>1</sup>, Maxime Derex<sup>1,4</sup>

<sup>1</sup> Institute for Advanced Study in Toulouse

<sup>2</sup> University of California, Davis

<sup>3</sup> Sciences Po, Paris

<sup>4</sup>Toulouse School of Economics, CNRS, University of Toulouse Capitole.

\* Shared first authorship

## Abstract

Through cumulative cultural evolution, populations discover novel solutions that allow them to solve complex problems. However, populations also reliably develop and sustain practices that are costly yet appear to have no effect on their stated outcome, such as bloodletting, divination, and malicious sorcery. Here, we show that ineffective, costly solutions can develop and be maintained because of their subjective appeal. In a pre-registered transmission-chain experiment, participants were tasked with developing a solution to a problem over several trials, after which they passed instructions onto the subsequent generation. The visual appearance of the experimental task was manipulated to vary the extent to which cognitive biases could interfere with participants' expectations of efficacy, enabling a direct test of the effect of cognitive appeal on cumulative cultural evolution. Participants in the low-appeal treatment ( $n = 240$ ) progressively developed more efficient solutions across generations. Participants in the high-appeal treatment ( $n = 240$ ), in contrast, produced costlier solutions with no increase in efficiency over the course of the experiment. These results suggest that, even under conditions favorable to cumulative cultural evolution, a form of cognitive appeal can alone prevent the process and promote the emergence and cultural persistence of ineffective solutions.

## 37 **Introduction**

38 Cumulative cultural evolution appears critical for humans' success<sup>1-3</sup>. Through the  
39 accumulation of cultural innovations over many generations, populations have developed  
40 solutions that no single individual could have invented, allowing us to inhabit new  
41 environments and achieve feats unparalleled in the animal world, from space exploration to  
42 industrial-scale agriculture. Several experiments have shown that cumulative cultural evolution  
43 occurs through instrumental, payoff-biased social learning, which allows the gradual  
44 accumulation of beneficial modifications over time<sup>4-6</sup>. Participants tasked with solving a  
45 problem, such as building a basket to carry rice or designing a wheel to roll down a track,  
46 preferentially adopt and retain solutions that yield greater expected payoffs<sup>7,8</sup>. When  
47 successive generations of learners select both successful demonstrators and effective-seeming  
48 solutions, cumulative cultural evolution can drive the emergence of complex, highly efficient  
49 solutions<sup>6,8</sup>.

50         Theoretical and empirical work has shown that cumulative cultural evolution operates  
51 under a wide range of conditions, albeit with differing levels of effectiveness. When payoffs  
52 are immediate and accurate, payoff-biased social learning allows individuals to filter out less  
53 efficient solutions, gradually bringing populations closer to the optimal solution over time<sup>9,10</sup>.  
54 However, when payoffs are noisy or delayed, gradual improvements are slower, as individuals  
55 struggle to accurately assess the effectiveness of solutions<sup>11</sup>. Additionally, experimental work  
56 has demonstrated that populations tend to converge on less efficient solutions in environments  
57 that are misaligned with their prior assumptions<sup>12</sup>. These mitigating factors notwithstanding,  
58 cumulative cultural evolution remains powerful enough to foster the emergence of increasingly  
59 efficient solutions across diverse conditions, even when those solutions are counterintuitive  
60 and beyond individual understanding<sup>8</sup>.

61         However, while cumulative cultural evolution reliably promotes the emergence of  
62 increasingly efficient solutions in experimental settings, it also appears to produce and preserve  
63 ineffective practices in the real world. Humans around the world engage in practices that seem  
64 to have no effect on their stated outcome, such as divination<sup>13,14</sup>, rain magic<sup>15</sup>, malicious  
65 sorcery<sup>16</sup>, and bloodletting<sup>17</sup>. These practices are widespread and elaborate, requiring non-  
66 trivial investments of time and resources. Research at the intersection of cognitive science and  
67 anthropology suggests that these practices may develop because of systematic biases in how  
68 individuals evaluate solutions. These biases are examples of "content biases" that can influence  
69 cultural evolution<sup>18</sup>. For example, prior research showed that participants in the United States

70 and Brazil evaluated rituals as more effective when they involved features like repetition or  
71 specific times, despite the rituals having no effect on their intended outcome<sup>19</sup>.

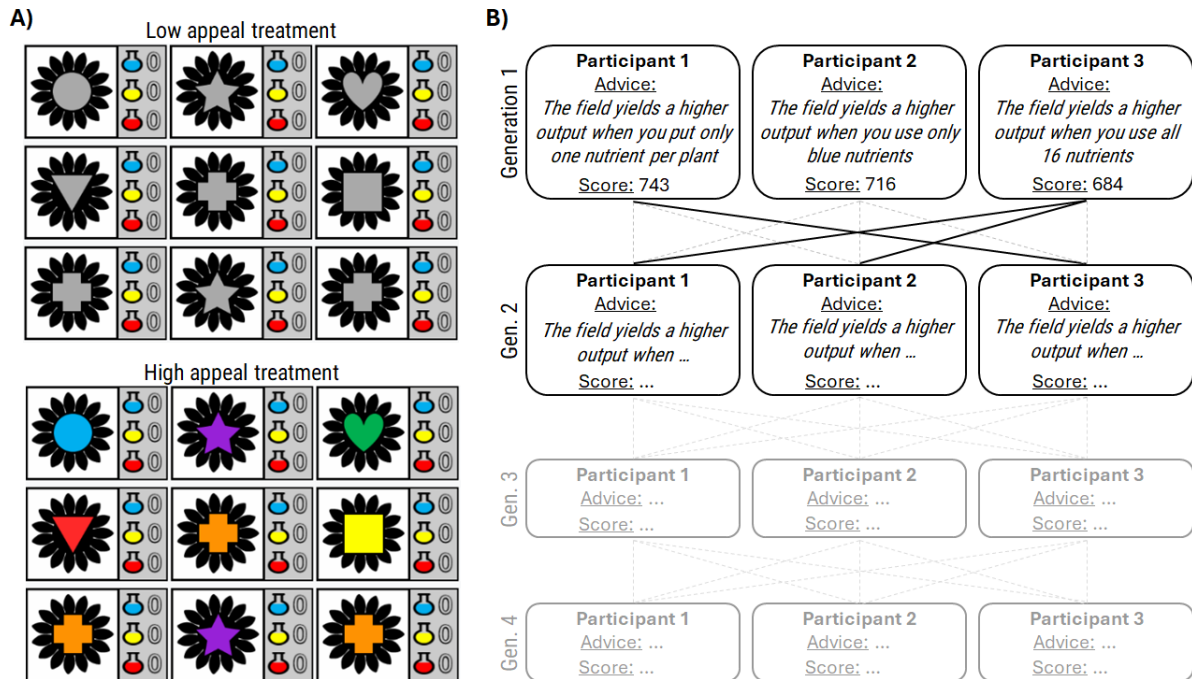
72 Of the features that can bias evaluations of potential solutions, one of the most  
73 widespread seems to be a visual correspondence between an intervention and its target.  
74 Anthropologists have long observed that people often use magic and folk medicines that  
75 visually resemble the desired outcome<sup>20</sup>. Well-known examples include voodoo dolls or  
76 effigies, in which a practitioner tries affecting another person by manipulating an object that  
77 resembles them, as well as the doctrine of signatures, which holds that herbs resembling  
78 specific body parts can be used to treat ailments affecting those parts<sup>21,22</sup>. Such  
79 correspondences have been hypothesized to tap into widespread cognitive biases<sup>23,24</sup>.

80 Here, we hypothesize that subjective appeal—in this case, the potential for similarity  
81 between an intervention and its target—can drive the emergence and persistence of ineffective  
82 solutions even in conditions otherwise conducive to cumulative cultural evolution. To test this  
83 hypothesis, we conducted a pre-registered experiment in which participants were tasked with  
84 maximizing their score by allocating nutrients to nine plants in a virtual field (Figure 1A).  
85 Nutrients came in three colors of different costs—blue (2 units), yellow (6 units), and red (10  
86 units)—and participants could allocate up to sixteen nutrients across the nine plants. This  
87 resulted in a vast solution space, with billions of distinct configurations. The field’s output  
88 varied as a function of how nutrients were allocated across plants, although scores were noisy  
89 through the inclusion of a random component (see Methods section). The optimal solution was  
90 to provide each plant with a single nutrient of the lowest cost (the blue nutrient).

91 Participants were organized into experimental chains of four generations, with each  
92 generation consisting of three participants (Figure 1B). Each participant completed six trials  
93 and received feedback on their score after each one. After their final trial, participants wrote a  
94 piece of advice for the next generation. From the second generation onward, participants were  
95 shown the final scores and written advice of three individuals from the previous generation and  
96 were asked to choose one as their cultural demonstrator before beginning their first trial.

97 We implemented two experimental treatments that differed in the degree of visual  
98 similarity between the features of the plants and those of the nutrients. In the low-appeal  
99 treatment (n=240), the plants had different shapes but were all gray in color (Figure 1A). In the  
100 high-appeal treatment (n=240), in contrast, plants were additionally assigned one of six colors:  
101 plant colors were either primary colors, matching the colors of the nutrients (blue, yellow, red),

102 or secondary colors that can typically be created by mixing the primary colors (green, purple,  
 103 orange). Thus, in this treatment, the nutrients fed to the plants could resemble the plants  
 104 themselves, potentially allowing for more visually appealing solutions than in the low-appeal  
 105 treatment. Importantly, the size of the solution space was kept equivalent across experimental  
 106 treatments by pairing each plant shape with a unique color (Figure 1A).



107  
 108 **Figure 1. Problem-Solving Task and Experimental Design.** A) Plant-Feeding Task: Participants  
 109 were presented with six uniquely shaped flowers within a field of nine plants. The plants appeared either  
 110 uniformly gray (low-appeal treatment) or in various colors, with each shape paired to a specific color  
 111 (high-appeal treatment). Participants had access to three different nutrient types and could use up to 16  
 112 nutrients across the entire field. After each trial, participants received feedback on the score their  
 113 solution generated. B) Participants were organized into experimental chains consisting of four  
 114 generations, with three participants per generation. After engaging with the task for six trials, each  
 115 participant wrote a piece of advice for participants in the subsequent generation. From Generation 2  
 116 onward, participants could select one cultural demonstrator based on that individual's written advice  
 117 and score in their final trial before solving the task and producing their own piece of advice for  
 118 participants in the following generation.

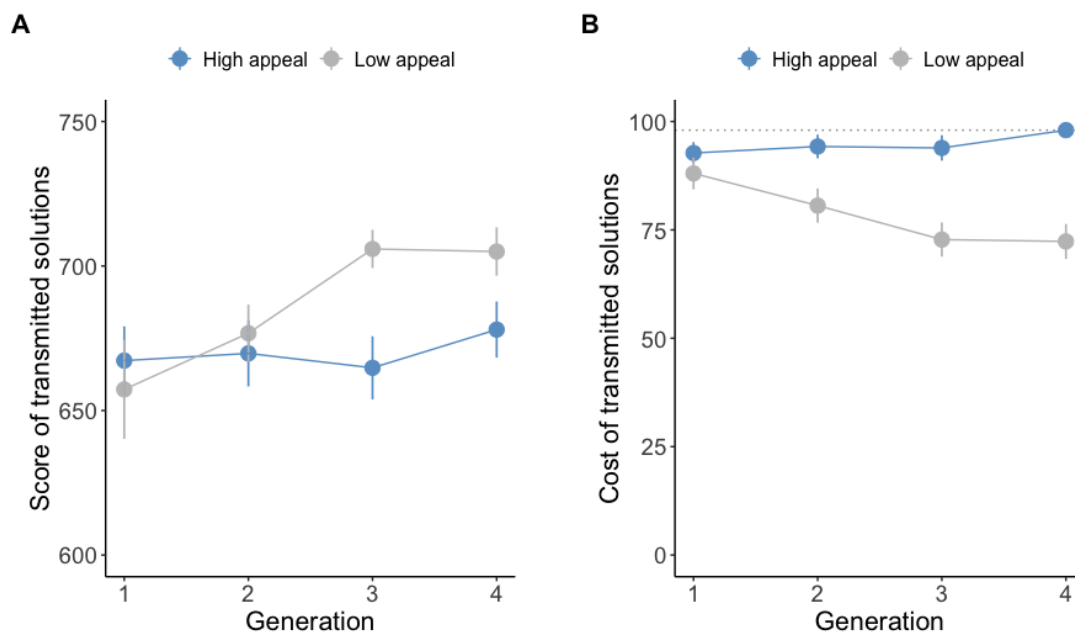
119  
 120 Under these conditions, rates of cumulative cultural evolution should be comparable across  
 121 treatments if cognitive biases do not influence participants' expectations of efficacy.  
 122 Conversely, if changes in the visual appearance of stimuli interfere with these expectations, we  
 123 would expect rates of cumulative cultural evolution to differ between treatments. To test this,  
 124 we compared the score and cost of culturally transmitted solutions—that is, the solutions  
 125 produced on the sixth trial—across generations in the low-appeal and high-appeal treatments.  
 126 We predicted that the score of transmitted solutions would increase over generations in the

127 low-appeal treatment and that this increase would exceed any change observed in the high-  
128 appeal treatment ([link to the preregistration](#)). Additionally, we predicted that participants in the  
129 low-appeal treatment would show stronger payoff-biased social learning when selecting their  
130 cultural demonstrator than participants in the high-appeal treatment.

131

## 132 Results

133 In the low-appeal treatment, the score of the transmitted solutions increased over the four  
134 generations (Figure 2A), while the cost of these solutions progressively decreased (Figure 2B,  
135 see Methods for details on score and cost calculation). In the high-appeal treatment, in contrast,  
136 neither the score of the transmitted solutions (Figure 2A) nor the cost of these solutions (Figure  
137 2B) changed significantly over the course of the experiment. In the low-appeal treatment,  
138 scores increased by an estimated 17.23 points with each additional generation ( $SE = 4.95$ ,  $p <$   
139  $0.001$ ), whereas in the high-appeal treatment, the change in score across generations was  
140 statistically indistinguishable from zero ( $\beta=2.72$ ,  $SE = 4.85$ ; Supplementary Table 1A).



141

142 **Figure 2. Score and cost of culturally transmitted solutions across generations.** A) In the low-  
143 appeal treatment (gray line), participants showed gradual improvements in solution scores across  
144 generations, whereas scores in the high-appeal treatment (blue line) did not change significantly. B)  
145 Participants in the low-appeal treatment progressively reduced the cost of their solutions, while those  
146 in the high-appeal treatment showed no statistically significant change over time. The dotted line  
147 represents the cost of the mix-and-match solution (see Figure 3 for an illustration).

148

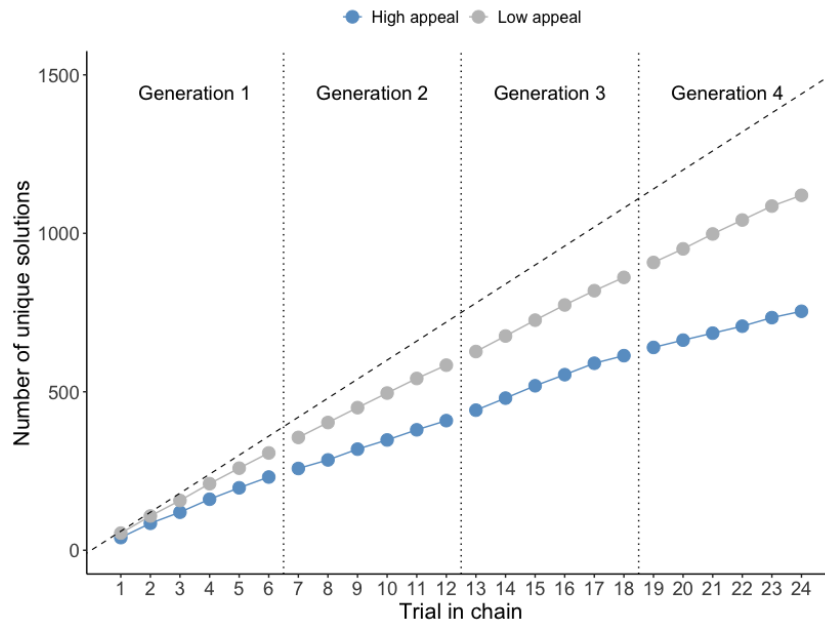
149 Among the three variables that affected solution scores (number of plants treated,  
150 number of nutrients used, and the type of nutrients used; see Methods), only the type of

151 nutrients differed between treatments, with the number of blue nutrients increasing across  
152 generations in the low-appeal treatment (Generation:  $\beta = -0.11$ , SE = 0.17,  $p = \text{ns}$ ; Low appeal:  
153  $\beta = -0.54$ , SE = 0.87,  $p = \text{ns}$ ; Generation  $\times$  Low appeal:  $\beta = 0.95$ , SE = 0.25,  $p < 0.001$ ; see  
154 Supplementary Figure A1 and Table A2 for details).

155 Contrary to our second prediction, we observed no differences between treatments in social  
156 learning strategies. Participants in both treatments relied on payoff-biased social learning when  
157 choosing among the three available cultural models, selecting the highest-scoring  
158 demonstrators 72% of the time in the low-appeal treatment and 78% of the time in the high-  
159 appeal treatment. The mean agreement among learners, calculated as the pairwise proportion  
160 of learners within a chain who selected the same demonstrator, was also comparable between  
161 treatments (67.2% and 66.1%, respectively; see Supplementary Figure A2), as was the  
162 probability of a participant implementing the inherited solution in their first trial (14.4% and  
163 18.3%, respectively; see Supplementary Figure A3.a and Table A3). However, the similarity  
164 between the solution participants inherited and the solution they used in their first trial was  
165 higher in the high-appeal treatment than in the low-appeal treatment overall (Generation:  $\beta =$   
166  $0.47$ , SE = 0.66,  $p = 0.48$ ; Low appeal:  $\beta = 6.61$ , SE = 3.00,  $p = 0.028$ ; Generation  $\times$  Low  
167 appeal:  $\beta = -1.45$ , SE = 0.93,  $p = 0.118$ ; see Supplementary Figure A3.b and Table A3 for  
168 details).

169 To understand why cumulative performance differed between treatments despite  
170 similar social learning strategies, we conducted exploratory analyses comparing exploration  
171 and cultural transmission patterns. Results from these analyses show that participants explored  
172 the solution space more extensively in the low-appeal treatment than in the high-appeal  
173 treatment. On average, they produced 5.43 unique solutions out of 6 trials in the low-appeal  
174 treatment, compared to 4.61 in the high-appeal treatment. Across all experimental replicates,  
175 77% of solutions were unique in the low-appeal treatment, compared to only 52% in the high-  
176 appeal treatment. Figure 3 shows the cumulative number of unique solutions across trials,  
177 illustrating that exploration was broader in the low-appeal condition. Reduced exploration was  
178 also apparent in analyses of within-participant solution diversity across generations, measured  
179 by the pairwise distance between solutions in six trials. Within-participant diversity tended to  
180 decrease across generations in both treatments, but remained overall lower in the high-appeal  
181 treatment (Generation:  $\beta = -1.59$ , SE = 0.34,  $p < 0.001$ ; Low appeal:  $\beta = 4.35$ , SE = 1.44,  $p =$   
182  $0.003$ ; Generation  $\times$  Low appeal:  $\beta = 0.05$ , SE = 0.48,  $p = 0.917$ ; see Supplementary Figure A4

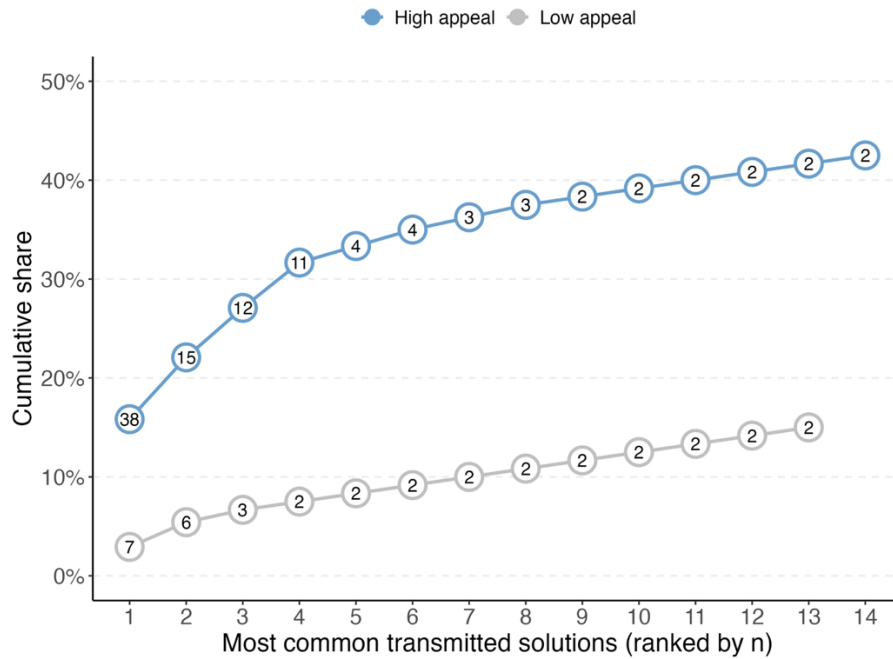
183 and Table A4 for details). Additionally, within-chain and between-chain diversity was lower  
184 in the high-appeal treatment than in the low-appeal treatment (see Supplementary Figure A5).  
185



186  
187 **Figure 3. Exploration across trials.** Cumulative number of unique solutions (i.e., nutrient-to-plant  
188 allocation patterns) across trials in each treatment. The dashed black line indicates the theoretical upper  
189 bound of exploration, assuming every participant submitted a solution not previously produced in this  
190 treatment. The gray line shows the cumulative number of unique solutions observed in the low-appeal  
191 treatment, while the blue line shows the high-appeal treatment.

192  
193 Analyses of the solutions produced in the sixth trial—that is, the solutions transmitted  
194 between generations—reveal important differences between treatments. In the high-appeal  
195 treatment, the most common solution accounted for about 16% of all transmitted solutions (38  
196 of 240, Figure 4), and the two most common solutions together accounted for about 22% of  
197 transmitted solutions (53 of 240). In the low-appeal treatment, no single solution was observed  
198 more than seven times, and the two most common solutions accounted for about 5% of all  
199 transmitted solutions (13 of 240). Overall, solutions observed more than once represented about  
200 41% of all transmitted solutions in the high-appeal treatment (99 of 240 solutions across 14  
201 unique solutions), compared with about 15% in the low-appeal treatment (36 of 240 solutions  
202 across 13 unique solutions).

203



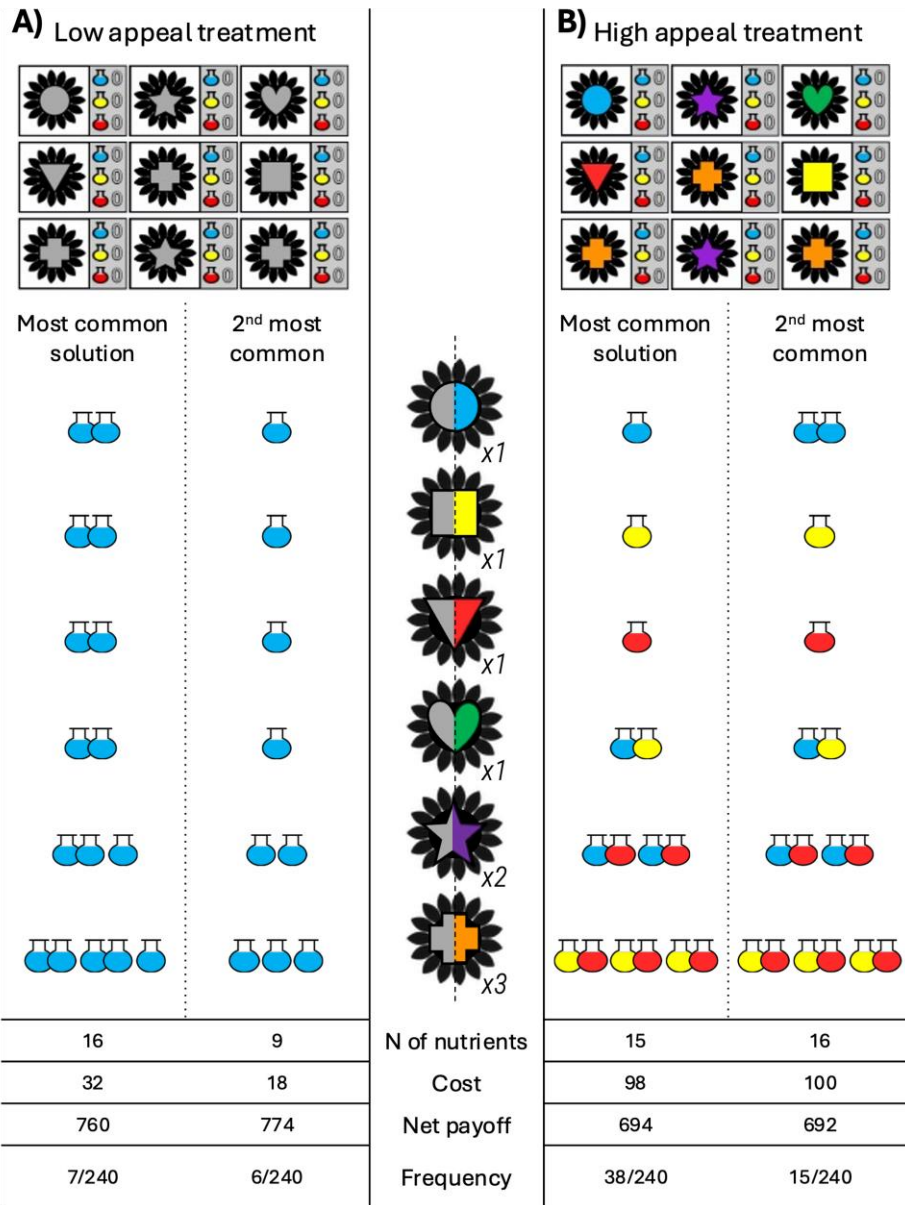
204

205 **Figure 4. Cumulative share of most common transmitted solutions.** Cumulative proportion of  
 206 transmitted solutions ranked by frequency in each treatment. In the high-appeal condition, a relatively  
 207 small number of solutions account for a substantial share of transmitted solutions. In the low-appeal  
 208 condition, transmitted solutions are more evenly distributed across a broader range of configurations,  
 209 with no single solution dominating to the same extent. Numbers inside the circles indicate the number  
 210 of times each solution was transmitted. Only solutions observed more than once are represented.

211

212 Figure 5 illustrates how the most common solutions visually differed between  
 213 treatments. In the low-appeal treatment, the two most common solutions involved only blue  
 214 nutrients: the most common used 16 blue nutrients, while the second most common—which  
 215 was the optimal solution—used 9 blue nutrients, one per plant (Figure 5). In the high-appeal  
 216 treatment, the most common solutions followed a costly 'mix-and-match' strategy: participants  
 217 matched nutrient colors to flower colors whenever possible, and when exact matches were  
 218 unavailable, they combined nutrient colors based on primary-color mixing (e.g., mixing blue  
 219 and red nutrients for a purple flower; Figure 5). These mix-and-match solutions cost  
 220 approximately five times more than the optimal solution.

221



222

223 **Figure 5. Most common culturally transmitted solutions.** The central column shows the composition  
 224 of the field, alongside the nutrient configurations used in the most common solutions for each treatment.  
 225 Note that the star/violet and orange/cross plants appeared twice and three times, respectively, in the  
 226 field. A) The two most common culturally transmitted solutions in the low-appeal treatment exclusively  
 227 involve blue nutrients. The most common solution in this treatment uses 16 blue nutrients, the most  
 228 efficient option when all nutrients are used. The second most common solution involves 9 blue  
 229 nutrients—one for each plant—which is the optimal solution. B) In the high-appeal treatment, the two  
 230 most common solutions use a mix-and-match strategy, where primary-colored plants are fed with the  
 231 corresponding colored nutrient, while secondary-colored plants receive a mix of colored nutrients.

232

233

234

235

236

Analyses of the written advice show that advice messages in the high-appeal treatment had higher semantic similarity, despite being longer, than those in the low-appeal treatment (see Supplementary Material and Supplementary Figure A7 for details). A classification of advice content indicates that participants in the high-appeal treatment transmitted advice

237 consistent with the transmitted feeding patterns, most often recommending the “mix-and-  
238 match” strategy throughout the experiment. By contrast, advice in the low-appeal treatment  
239 increasingly recommended favoring blue nutrients over generations, shifting from relatively  
240 uncommon in the first generation to being frequently transmitted by the end of the experiment  
241 (see Supplementary Figure A8).

242

## 243 **Discussion**

244 Our findings demonstrate that changes to the visual appearance of stimuli, by increasing the  
245 opportunity for cognitive biases to interfere with expectations of efficacy, can hinder  
246 cumulative cultural evolution. In the low-appeal treatment, where all plants were gray,  
247 solutions became progressively more efficient across generations. In contrast, in the high-  
248 appeal treatment, where plants were colorful, participants produced costly, ineffective  
249 solutions and showed no gradual improvement across generations.

250         This failure to shift toward more efficient solutions in the high-appeal treatment appears  
251 to be driven by reduced exploration both within and across participants. Our findings suggest  
252 that this lower level of exploration reflects the strong appeal of the “mix-and-match” strategy  
253 in the high-appeal treatment, wherein participants aligned nutrient colors with flower colors  
254 whenever possible or combined nutrient colors based on primary-color mixing. The appeal of  
255 this strategy is evident in several ways. First, participants in the high-appeal treatment produced  
256 fewer unique solutions, indicating a shift toward greater exploitation in the exploration–  
257 exploitation dilemma faced by participants. This pattern is consistent with the idea that  
258 cognitive biases shaped participants’ expectations of efficacy in our experiment. Second, when  
259 participants did explore, exploration was more local and more convergent in the high-appeal  
260 treatment: successive solutions produced by the same participant were more similar (smaller  
261 within-individual solution distances across the six trials), and solutions overlapped more across  
262 participants (smaller within-chain and between-chain solution distances).

263         Limited exploration may also have been self-reinforcing through the cultural  
264 transmission of advice. Experiments have shown that receiving a theory can channel  
265 exploration toward a specific region of the search space<sup>8,25</sup>. In our experiment, the limited  
266 variation in the theories transmitted in the high-appeal treatment, which tended to focus on the  
267 mix-and-match solution, would therefore be expected to channel exploration toward a similar  
268 region across participants, further constraining the total exploration across individuals.

269 Regardless of the precise mechanism, the observed reduction in variation constrains cultural  
270 evolvability—the capacity of a cultural system to generate, retain, and build upon culturally  
271 transmitted variation—by limiting the raw material on which selection operates<sup>26</sup>.

272 Importantly, the absence of gradual improvement across generations in the high-appeal  
273 treatment occurred even though participants relied on payoff-biased social learning, most often  
274 selecting the highest-scoring cultural demonstrators. Payoff-biased social learning is known to  
275 be critical for the emergence of increasingly efficient solutions, which is consistent our  
276 previous version of the same experiment in which payoff information was unavailable for  
277 demonstrator choice and no cumulative improvement was observed in either treatment (see  
278 Supplementary Material). Disrupting payoff-biased social learning would therefore be  
279 expected to hinder cumulative cultural evolution even more than what we observe here.

280 A likely reason that participants continued to rely on payoff information, despite the  
281 opportunity to also judge demonstrators by the appeal of their advice, is that payoff comparison  
282 was cognitively straightforward (participants compared three numerical scores), whereas  
283 comparing advice required reading and integrating several text messages. Outside the  
284 laboratory, payoff information is often noisier, delayed, or difficult to quantify, which could  
285 increase reliance on appeal. A second factor is that strong convergence in the high-appeal  
286 condition limited variation in the appeal of competing messages. Consistent with this, advice  
287 messages in the high-appeal condition were more homogeneous than in the low-appeal  
288 condition and thus less diagnostic for differentiating demonstrators, making payoffs a more  
289 informative cue.

290 The core feature of mix-and-match—a perceptual alignment between the intervention  
291 and the target object—mirrors patterns found in magic and folk remedies, exemplifying what  
292 anthropologists term the “law of similarity”<sup>20</sup>. Such cultural patterns suggest that humans  
293 reliably develop tendencies to infer causal relationships from visual similarity. The reasons for  
294 such causal biases remain debated; proposals include that people erroneously learn causal  
295 relationships from environmental regularities<sup>23</sup> and that the principle of like-affects-like occurs  
296 frequently enough in the natural world that it proved a useful scaffold for understanding  
297 causality for our ancestors<sup>21</sup>. Whatever their developmental or evolutionary origins, such  
298 intuitions can lead individuals to overestimate causal links between features and effects, such  
299 as assuming that a plant’s shape is linked to its medicinal properties or, as in our experiment,  
300 that nutrients are more effective when their colors match those of the plants to which they are

301 applied. The exposure to cultural products that harness these intuitions, such as video games,  
302 may further amplify such causal associations.

303         Given the apparent functionality of many cultural practices found across much of  
304 culture, researchers have often explained those that seem instrumentally ineffective, such as  
305 shamanism, magic, and divination, by proposing alternative functional benefits, such as  
306 enhanced confidence or social cohesion <sup>27,28,29</sup>. However, consistent with recent arguments  
307 <sup>13,30,31</sup>, our results experimentally illustrate that ineffective cultural practices can be maintained  
308 by subjective appeal. Our experiment investigated a single form of subjective appeal—visual  
309 similarity—and the resulting cultural product was much simpler than practices like divination  
310 or shamanism. We suspect that these more complex practices tap into other dimensions of  
311 subjective appeal beyond those investigated here. Nevertheless, our study serves as proof of  
312 concept, demonstrating that functional benefits are not required for the emergence and  
313 persistence of costly ineffective cultural practices.

314         Our study has important implications for studying cumulative cultural evolution under  
315 controlled conditions. Although researchers have investigated the conditions that constrain  
316 cumulative cultural evolution <sup>5,12,32-37</sup>, most experiments are designed in ways that may  
317 artificially inflate the potential for gradual cultural improvement. For example, with few  
318 exceptions <sup>11</sup>, researchers typically provide participants with immediate and accurate payoffs,  
319 despite many interventions in the real-world leading to payoffs that are noisier and time-  
320 delayed. More importantly, the solution space in laboratory experiments is often artificially  
321 constrained, limiting the opportunity for inefficient but cognitively appealing solutions to  
322 compete with more efficient ones. These design choices potentially lead to an  
323 underappreciation of the inefficiencies of cumulative cultural evolution as manifested both in  
324 practices that are cheap and effective yet rare, such as handwashing <sup>38</sup>, and in practices that are  
325 common and costly yet futile.

326         In our experiment, the high-appeal treatment created opportunities for visual similarity,  
327 which we expected would bias how participants evaluated potential solutions. By embedding  
328 this cognitive bias into the task environment, we created a setting in which visually appealing—  
329 but costly and inefficient—solutions could plausibly emerge. This does not mean, however,  
330 that the experiment was artificially tailored to produce such outcomes. First, the patterns  
331 observed in the low-appeal treatment confirm that the conditions of the experiment—such as  
332 task complexity, payoff noise, and number of demonstrators—were conducive to cumulative  
333 cultural evolution. Second, even in the high-appeal treatment, the opportunity for cognitively

334 appealing solutions to emerge was limited. In real-world contexts, learners routinely encounter  
335 similarity-based cues alongside many other sources of biased inference. Outside the lab,  
336 individuals interact with a broader set of variables—such as timing and other contextual  
337 factors—that may trigger additional biases not captured in our design<sup>19</sup>. This broader and more  
338 complex inferential landscape may further amplify the emergence and persistence of  
339 ineffective solutions, underscoring the importance of understanding how subjective appeal  
340 shapes the dynamics of cultural evolution across contexts and domains.

341

## 342 **Methods**

343 **Ethics Statement.** All methods were approved by the Toulouse School of Economics/Institute  
344 for Advanced Studies in Toulouse Review Board for Ethical Standards in Research (2022-11-  
345 001). All participants provided informed consent before taking part in the experiment. All  
346 individual information has been anonymized.

347 **Participants.** A total of 480 participants were recruited for the study. Participants were  
348 recruited through Prolific and filtered based on several criteria: they were required to be aged  
349 between 18 and 35 years, from the United Kingdom, United States, Ireland, Australia, Canada,  
350 or New Zealand, and to have English as their first language. Additionally, participants needed  
351 to have no issues seeing colors and maintain an approval rate of 98-100% on Prolific.  
352 Participants received a £1.2 participation fee, with an average bonus of £1.8 based on their  
353 performance.

354 **Experimental task.** The experiment relied on a plant-feeding task in which participants were  
355 presented with a virtual field containing nine plants (Figure 1A). Participants aimed to  
356 maximize the output of their virtual field across 6 trials by feeding the plants with three types  
357 of nutrients that varied in their cost: blue (2 units), yellow (6 units), and red (10 units).  
358 Participants' decisions in each trial of nutrient allocation would impact the output of their field  
359 and the cost of their solution, which affected their score for that trial. The field's output was  
360 determined by the number of plants that receive at least one nutrient and a random component.  
361 Additional nutrients on the same plant, the shapes and colors of plants, or the color of the  
362 nutrients used had no effect on the output. Put precisely, the field's output was the number of  
363 plants fed, multiplied by a uniformly distributed number between 60 and 100. Subtracting the  
364 total nutrient cost and adding a fixed score of 72 resulted in the participant's score on that trial:

365 
$$\text{Score} = 72 + \text{Uniform}(60, 100) \times \text{Number of plants fed} - \text{Total cost of nutrients}$$

366           The fixed score of 72 points was added to prevent participants from having a negative  
367 score in the worst-case scenario.

368           Participants were limited to using a maximum of 16 nutrients across the entire field,  
369 with each plant receiving no more than 9 nutrients of a given color. Under these constraints,  
370 trial scores could range from 0 to 954. The optimal strategy (feeding each plant a single blue  
371 nutrient) yielded an average score of 774. However, in the best-case scenario, the field could  
372 produce an output of 900 (9 plants  $\times$  100), at a cost of 18, resulting in a total score of  $72 + 900$   
373  $- 18 = 954$ . In contrast, the worst strategy (feeding a single plant 9 red and 7 yellow nutrients)  
374 yielded an average score of 20. In the worst-case scenario, this approach could produce a field  
375 output of just 60 at a cost of 132, resulting in a total score of  $72 + 60 - 132 = 0$ .

376           Participants were informed about the random component affecting their score. At the  
377 end of each trial, they were shown their score for that trial. However, they were not told the  
378 exact function governing the field's output and the costs of the nutrients.

379 **Treatments.** The study implemented two treatments: low-appeal and high-appeal (see Figure  
380 1A). A total of 480 participants were equally and randomly assigned to one of the two treatment  
381 conditions: high-appeal ( $n = 240$ ) and low-appeal ( $n = 240$ ). Each condition consisted of 20  
382 transmission chains, each comprising four generations of three participants, for a total of 12  
383 participants per chain. In the low-appeal treatment, all plants were gray and exhibited one of  
384 six distinct shapes: circle, star, heart, triangle, cross, and square (see Figure 1A, upper figure).  
385 The high-appeal treatment featured the same shapes but in distinct colors: the round shape was  
386 blue, the star was purple, the heart was green, the triangle was red, the cross was orange, and  
387 the square was yellow (see Figure 1A, lower figure). All participants in each treatment saw the  
388 same fixed arrangement of the plants. In both treatments, the shapes and colors of the plants  
389 were purely aesthetic differences and did not influence the output of the field.

390 **Cultural transmission.** The participants engaged with the task over six trials. Each  
391 participant's last solution and its corresponding score for that trial were passed onto the next  
392 participant in their chain. In addition to transmitting their solutions, each participant was  
393 required to provide an accompanying piece of advice for the next participant in their chain (see  
394 Figure 1B). The advice was pre-formatted to begin with the phrase "*The field yields a higher*  
395 *output when...*" to ensure comparability across responses. Before starting the task, participants  
396 were provided with three pieces of advice from participants in the previous generation, along  
397 with their score from the last trial, and could select one participant as cultural demonstrator.

398 Upon selecting a cultural demonstrator, participants were provided with the chosen piece of  
399 advice along with the participants' solution (i.e., which nutrients were provided to each plant)  
400 for the duration of the experiment.

#### 401 **Statistical analysis.**

402 *Score and cost of transmitted solutions.* We used the transmitted solutions of each of the 480  
403 participants as the unit of analysis in our models. The response variable was each participant's  
404 expected score—that is, the score excluding the effect of the random component—to reduce  
405 the impact of artificial noise on the statistical analyses. To compare how the score changed  
406 over the course of the experiment across the two treatment conditions, we first conducted two  
407 separate models using subsets of the data for each treatment ( $n = 240$  for each,  $n_{\text{chain}} = 20$ ). To  
408 assess how the expected score changed across generations within a given treatment condition,  
409 we used a multilevel regression model with an interaction term between generation and  
410 treatment as a fixed effect, with the pooled data (high-appeal:  $n = 240$ , low-appeal:  $n = 240$ ,  
411  $n_{\text{chain}} = 40$ ). We repeated those three models using participants' costs as the response variable.  
412 In all models, we accounted for chain-specific random effects.

413 *Social learning strategies.* The response variable was a binary indicator of whether participants  
414 selected the highest-scoring demonstrator among the three available cultural models. We used  
415 a multilevel model with treatment as a fixed effect and included chain and generation as random  
416 effects.

417 All data processing was performed using the R software environment<sup>39</sup>, and we conducted  
418 multilevel regression analyses with the `lmer` function from the `lme4` package in R<sup>40</sup>.

419 **Pre-registration.** The experiment was pre-registered on AsPredicted prior to data collection  
420 ([link](#)).

421 **Data and Code Availability.** The data and scripts for the analyses are publicly available at:  
422 <https://github.com/aseyq/inefficient-solutions>

#### 423 424 **References**

- 425 1 Henrich, J. *The secret of our success: how culture is driving human evolution,*  
426 *domesticating our species, and making us smarter.* (Princeton University Press,  
427 2015).
- 428 2 Boyd, R., Richerson, P. J. & Henrich, J. The cultural niche: Why social learning is  
429 essential for human adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10918-10925  
430 (2011).

- 431 3 Derex, M. Human cumulative culture and the exploitation of natural phenomena.  
432 *Philosophical Transactions of the Royal Society B: Biological Sciences* **377**,  
433 20200311 (2021). <https://doi.org/10.1098/rstb.2020.0311>
- 434 4 Mesoudi, A. An experimental comparison of human social learning strategies: payoff-  
435 biased social learning is adaptive but underused. *Evol. Hum. Behav.* **32**, 334-342  
436 (2011).
- 437 5 Derex, M., Beugin, M.-P., Godelle, B. & Raymond, M. Experimental evidence for the  
438 influence of group size on cultural complexity. *Nature* **503**, 389-391 (2013).  
439 <https://doi.org/10.1038/nature12774>
- 440 6 Thompson, B., Opheusden, B. v., Sumers, T. & Griffiths, T. L. Complex cognitive  
441 algorithms preserved by selective social learning in experimental populations. *Science*  
442 **376**, 95-98 (2022). <https://doi.org/doi:10.1126/science.abn0915>
- 443 7 Zwirner, E. & Thornton, A. Cognitive requirements of cumulative culture: teaching is  
444 useful but not essential. *Scientific Reports* **5**, 16781 (2015).  
445 <https://doi.org/10.1038/srep16781>  
446 <http://www.nature.com/articles/srep16781#supplementary-information>
- 447 8 Derex, M., Bonnefon, J.-F., Boyd, R. & Mesoudi, A. Causal understanding is not  
448 necessary for the improvement of culturally evolving technology. *Nature Human*  
449 *Behaviour* **3**, 446-452 (2019). <https://doi.org/10.1038/s41562-019-0567-9>
- 450 9 Henrich, J. Demography and cultural evolution: How adaptive cultural processes can  
451 produce maladaptive losses - The Tasmanian case. *Am. Antiq.* **69**, 197-214 (2004).
- 452 10 Lehmann, L. & Wakano, J. Y. The handaxe and the microscope: individual and social  
453 learning in a multidimensional model of adaptation. *Evol. Hum. Behav.* **34**, 109-117  
454 (2013).
- 455 11 Caldwell, C. A. & Millen, A. E. Conservatism in laboratory microsocieties:  
456 unpredictable payoffs accentuate group-specific traditions. *Evol. Hum. Behav.* **31**,  
457 123-130 (2010).
- 458 12 Thompson, B. & Griffiths, T. L. Human biases limit cumulative innovation.  
459 *Proceedings of the Royal Society B: Biological Sciences* **288**, 20202752 (2021).  
460 <https://doi.org/doi:10.1098/rspb.2020.2752>
- 461 13 Hong, Z. & Henrich, J. The Cultural Evolution of Epistemic Practices. *Human Nature*  
462 **32**, 622-651 (2021). <https://doi.org/10.1007/s12110-021-09408-6>
- 463 14 Boyer, P. Why Divination?: Evolved Psychology and Strategic Interaction in the  
464 Production of Truth. *Curr. Anthropol.* **61**, 100-123 (2020).  
465 <https://doi.org/10.1086/706879>
- 466 15 Hong, Z., Slingerland, E. & Henrich, J. Magic and Empiricism in Early Chinese  
467 Rainmaking: A Cultural Evolutionary Analysis. *Curr. Anthropol.* **65**, 343-363 (2024).  
468 <https://doi.org/10.1086/729118>
- 469 16 Singh, M. Magic, Explanations, and Evil: The Origins and Design of Witches and  
470 Sorcerers. *Curr. Anthropol.* **62**, 2-29 (2021). <https://doi.org/10.1086/713111>
- 471 17 Miton, H., Claidière, N. & Mercier, H. Universal cognitive mechanisms explain the  
472 cultural success of bloodletting. *Evol. Hum. Behav.* **36**, 303-312 (2015).  
473 <https://doi.org/10.1016/j.evolhumbehav.2015.01.003>
- 474 18 Henrich, J. & McElreath, R. The evolution of cultural evolution. *Evol. Anthropol.* **12**,  
475 123-135 (2003).
- 476 19 Legare, C. H. & Souza, A. L. Evaluating ritual efficacy: Evidence from the  
477 supernatural. *Cognition* **124**, 1-15 (2012).  
478 <https://doi.org/https://doi.org/10.1016/j.cognition.2012.03.004>
- 479 20 Frazer, J. G. & Fraser, R. *The Golden Bough: A Study in Magic and Religion*.  
480 (Macmillan, 1890).

- 481 21 Rozin, P. & Nemeroff, C. in *Heuristics and Biases: The Psychology of Intuitive*  
482 *Judgment* (eds Thomas Gilovich, Dale Griffin, & Daniel Kahneman) 201-216  
483 (Cambridge University Press, 2002).
- 484 22 Bennett, B. C. Doctrine of Signatures: An Explanation of Medicinal Plant Discovery  
485 or Dissemination of Knowledge? *Economic Botany* **61**, 246-255 (2007).
- 486 23 Hong, Z. A cognitive account of manipulative sympathetic magic. *Religion, Brain &*  
487 *Behavior* **12**, 254-270 (2022). <https://doi.org/10.1080/2153599X.2021.2006294>
- 488 24 Nemeroff, C. & Rozin, P. in *Imagining the Impossible: Magical, Scientific, and*  
489 *Religious Thinking in Children* (eds Karl S. Rosengren, Carl N. Johnson, & Paul L.  
490 Harris) 1-34 (Cambridge University Press, 2000).
- 491 25 Derex, M., Bonnefon, J.-F., Boyd, R., McElreath, R. & Mesoudi, A. Social learning  
492 preserves both useful and useless theories by canalizing learners' exploration.  
493 *Proceedings of the Royal Society B: Biological Sciences* **292**, 20242499 (2025).  
494 <https://doi.org/doi:10.1098/rspb.2024.2499>
- 495 26 Schimmelpfennig, R., Razek, L., Schnell, E. & Muthukrishna, M. Paradox of  
496 diversity in the collective brain. *Philosophical Transactions of the Royal Society B:*  
497 *Biological Sciences* **377**, 20200316 (2022). <https://doi.org/doi:10.1098/rstb.2020.0316>
- 498 27 Damisch, L., Stoberock, B. & Mussweiler, T. Keep Your Fingers Crossed! How  
499 Superstition Improves Performance. *Psychol. Sci.* **21**, 1014-1020 (2010).
- 500 28 Watson-Jones, R. E. & Legare, C. H. The social functions of shamanism. *Behav.*  
501 *Brain Sci.* **41**, e88 (2018). <https://doi.org/10.1017/S0140525X17002199>
- 502 29 Moore, O. K. Divination - A New Perspective. *Am. Anthropol.* **59**, 69-74 (1957).
- 503 30 Singh, M. Subjective selection and the evolution of complex culture. *Evolutionary*  
504 *Anthropology: Issues, News, and Reviews* **31**, 266-280 (2022).  
505 <https://doi.org/https://doi.org/10.1002/evan.21948>
- 506 31 Singh, M. The cultural evolution of shamanism. *Behav. Brain Sci.* **41**, e66 (2018).  
507 <https://doi.org/10.1017/S0140525X17001893>
- 508 32 Griffiths, T. L., Kalish, M. L. & Lewandowsky, S. Theoretical and empirical evidence  
509 for the impact of inductive biases on cultural evolution. *Philosophical Transactions of*  
510 *the Royal Society B: Biological Sciences* **363**, 3503-3514 (2008).  
511 <https://doi.org/10.1098/rstb.2008.0146>
- 512 33 Schillinger, K., Mesoudi, A. & Lycett, S. J. Copying error and the cultural evolution  
513 of "additive" vs. "reductive" material traditions: an experimental assessment. *Am.*  
514 *Antiq.* **79**, 128-143 (2014).
- 515 34 Schillinger, K., Mesoudi, A. & Lycett, S. J. The impact of imitative versus emulative  
516 learning mechanisms on artifactual variation: implications for the evolution of  
517 material culture. *Evol. Hum. Behav.* (2015).  
518 <https://doi.org/http://dx.doi.org/10.1016/j.evolhumbehav.2015.04.003>
- 519 35 Mesoudi, A. Variable Cultural Acquisition Costs Constrain Cumulative Cultural  
520 Evolution. *PLoS One* **6**, e18239 (2011).
- 521 36 Derex, M., Edmiston, P., Lupyan, G. & Mesoudi, A. Trade-offs, control conditions,  
522 and alternative designs in the experimental study of cultural evolution. *Proceedings of*  
523 *the National Academy of Sciences* **121**, e2322886121 (2024).  
524 <https://doi.org/doi:10.1073/pnas.2322886121>
- 525 37 Caldwell, C. A., Schillinger, K., Evans, C. L. & Hopper, L. M. End State Copying by  
526 Humans (*Homo sapiens*): Implications for a Comparative Perspective on Cumulative  
527 Culture. *J. Comp. Psychol.* **126**, 161-169 (2012).
- 528 38 Bloomfield, S. F. & Nath, K. J. in *Somerset, England: International Scientific Forum*  
529 *on Home Hygiene*.

530 39 R: A Language and Environment for Statistical Computing (R Foundation for  
531 Statistical Computing, Vienna, Austria, 2011).  
532 40 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models  
533 Using lme4. *Journal of Statistical Software* **67**, 1 - 48 (2015).  
534 <https://doi.org/10.18637/jss.v067.i01>  
535  
536  
537

538

## Supplementary material

### 539 Preregistered Results and Additional Analyses

A) Scores			
	Low-Appeal only	High-Appeal only	Pooled
(Intercept)	643.17 *** (13.94)	663.15 *** (13.60)	663.15 *** (13.77)
Generation	17.23 *** (4.95)	2.72 (4.85)	2.72 (4.90)
Low Appeal			-19.98 (19.48)
Generation x Low Appeal			14.51 * (6.93)
B) Costs			
(Intercept)	92.17 *** (5.40)	90.85 *** (3.09)	90.85 *** (4.40)
Generation	-5.49 *** (1.50)	1.55 (1.12)	1.55 (1.32)
Low Appeal			1.32 (6.22)
Generation x Low Appeal			-7.04 *** (1.87)
Num. obs.	240	240	480
Num. groups: chain	20	20	40

540

541

542 **Table A1. Multilevel model estimates for expected scores (A) and costs (B) of culturally transmitted solutions**

543 **across treatments.** Separate models are reported for the low-appeal condition, high-appeal condition, and pooled

544 data to compare changes in expected scores/costs over generations, with random intercepts controlling for

545 transmission *chain* structure of the experiment. Standard errors are shown in parentheses and statistical

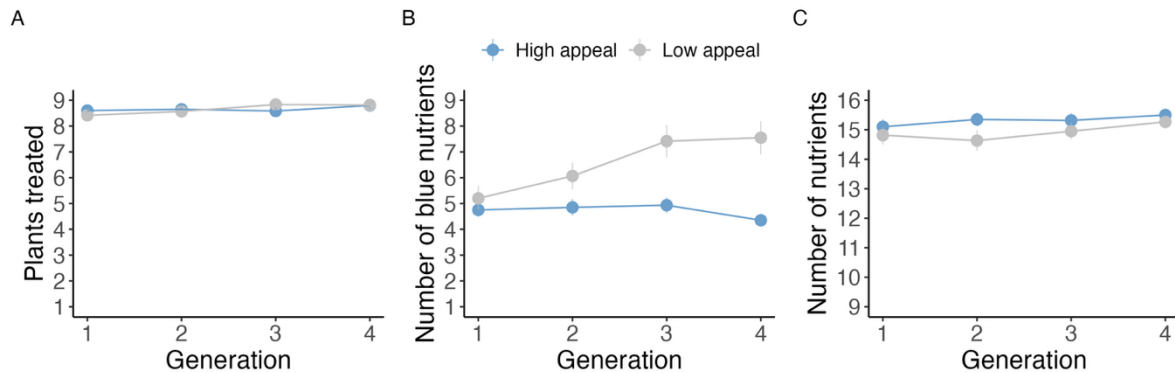
significance is denoted by asterisks (\*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$ ).

546

547

548

549 **Patterns of nutrient use**



550

551 **Figure A1. Patterns of nutrient use.** (A) Mean number of plants treated (i.e., plants receiving at least  
 552 one nutrient). (B) Mean number of blue nutrients used. (C) Mean total number of nutrients used. Error  
 553 bars represent ±1 SE. Across treatments, participants consistently fed most plants (A) and used most  
 554 available nutrients (C). However, in the low-appeal condition, the use of blue nutrients—the least costly  
 555 nutrient—increased across generations. In contrast, blue nutrient use remained stable across generations  
 556 in the high-appeal condition.

557

558

	Plants treated		Num. Blue Nutrients		Num. Nutrients	
<b>(Intercept)</b>	8.41 *** (0.13)	8.53 *** (0.16)	3.81*** (0.54)	5.00*** (0.62)	14.96*** (0.26)	15.02*** (0.33)
<b>Generation</b>	0.10* (0.04)	0.05 (0.06)	0.36** (0.12)	-0.11 (0.17)	0.14 (0.08)	0.12 (0.11)
<b>Low Appeal</b>	0.00 (0.09)	-0.23 (0.23)	1.84** (0.62)	-0.54 (0.87)	-0.40 (0.23)	-0.52 (0.46)
<b>Generation x Low Appeal</b>		0.09 (0.08)		0.95*** (0.25)		0.05 (0.16)
<b>Num. obs.</b>	480	480	480	480	480	480
<b>Num. groups: chain</b>	40	40	40	40	40	40

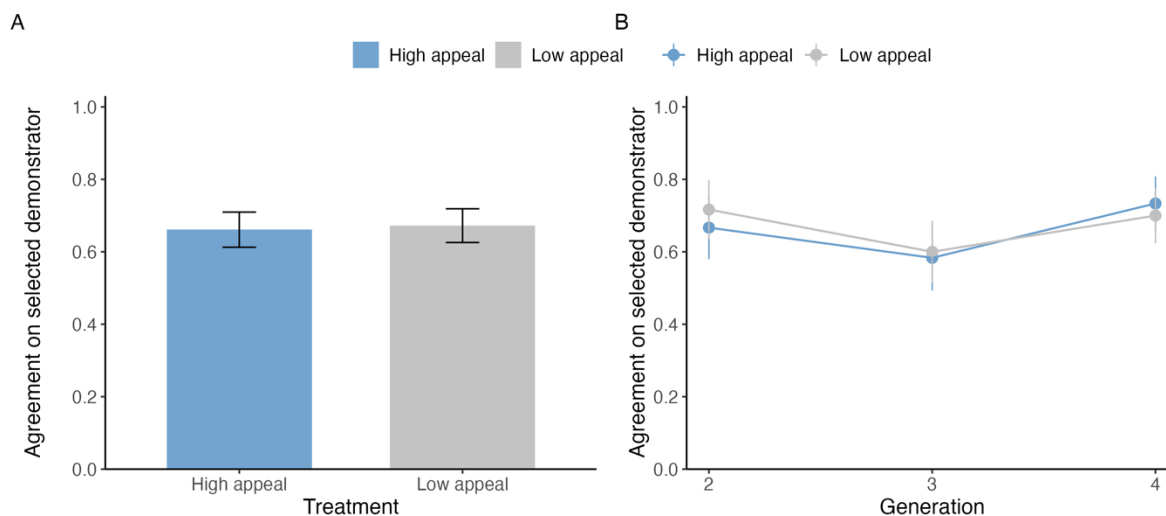
559 **Table A2. Multilevel model estimates for the number of plants treated, the number of blue nutrients used,**  
 560 **and the total number of nutrients used in transmitted solutions across treatments and generations.** Separate  
 561 models are reported for each dependent variable using pooled data. All models include random intercepts for  
 562 transmission chain to account for the hierarchical structure of the data. Standard errors are shown in parentheses  
 563 and statistical significance is denoted by asterisks (\*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$ ).

564

565

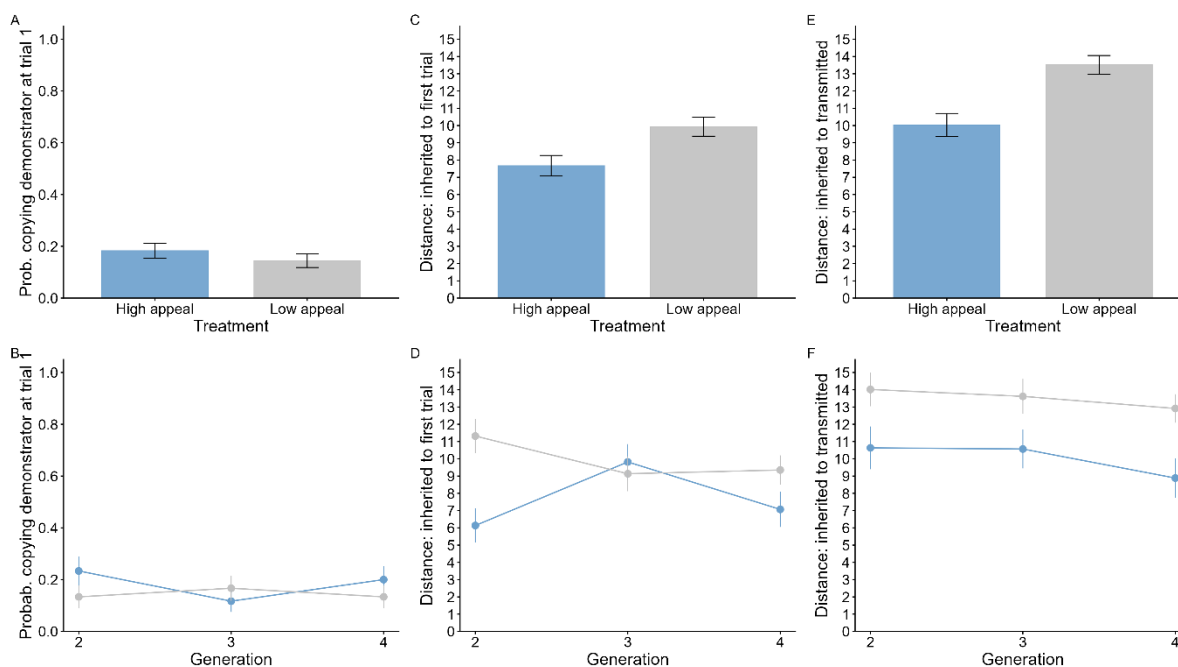
566

567



569 **Figure A2. Agreement among learners in demonstrator selection.** (A) Mean agreement within each treatment.  
 570 (B) Mean agreement by generation and treatment. Agreement is computed as the proportion of learner pairs within  
 571 a chain who chose the same demonstrator. Error bars represent ±1 SE. Levels of agreement are similar across  
 572 treatments and remain relatively stable across generations, indicating comparable patterns of demonstrator  
 573 selection in the high- and low-appeal conditions.  
 574  
 575

576 Social Influence



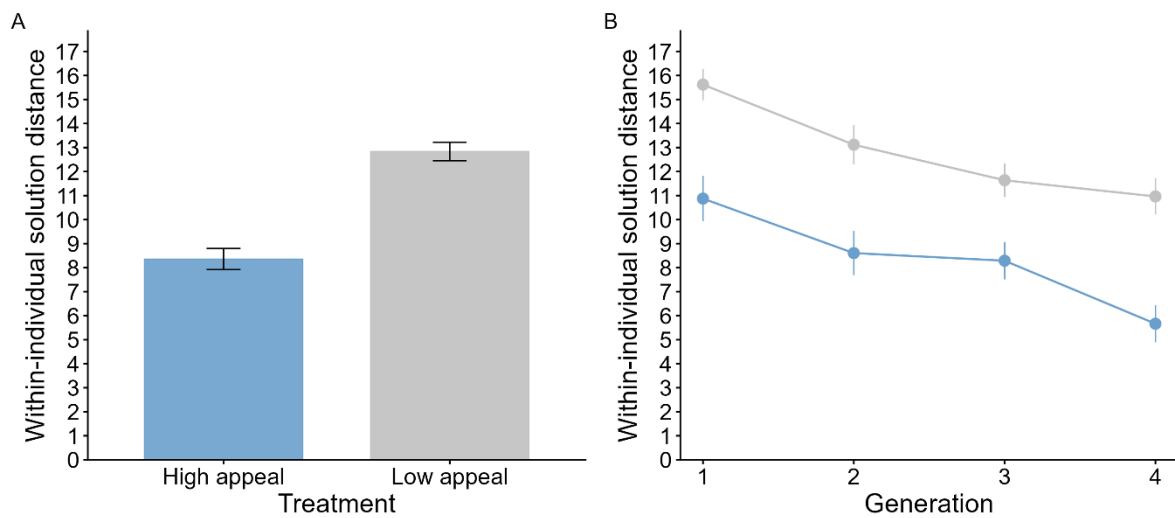
577 **Figure A3. Social influence.** (A–B) Probability of copying the inherited solution at Trial 1, defined as the  
 578 proportion of participants whose first-trial solution exactly matched the inherited solution from their selected  
 579 demonstrator. (C–D) Mean distance between the inherited solution and the solution used at Trial 1. (E–F) Mean  
 580 distance between the inherited solution and the transmitted solution (Trial 6). Distance is measured as the sum of  
 581 absolute differences across all nutrient allocations in the solution vector (Manhattan distance). Panels A, C, and  
 582 E show treatment-level means, whereas panels B, D, and F show means by treatment and generation. Error bars  
 583 represent ±1 SE.  
 584  
 585

	Probability of copying inherited solution at trial 1			Distance between inherited solution and solution at trial 1			Distance between inherited solution and transmitted solution		
<b>(Intercept)</b>	0.18*** (0.04)	0.21** (0.08)	0.23* (0.10)	7.67* ** (0.80)	8.45*** (1.61)	6.27** (2.12)	10.03*** (0.92)	12.17*** (1.71)	12.65*** (2.24)
<b>Low Appeal</b>	-0.04 (0.05)	-0.04 (0.05)	-0.09 (0.15)	2.26* (1.13)	2.26* (1.13)	6.61* (3.00)	3.49** (1.30)	3.49** (1.30)	2.51 (3.16)
<b>Generation</b>		-0.01 (0.02)	-0.02 (0.03)		-0.26 (0.47)	0.47 (0.66)		-0.71 (0.48)	-0.88 (0.68)
<b>Generation x Low Appeal</b>			0.02 (0.05)			-1.45 (0.93)			0.33 (0.96)
<b>Num. obs.</b>	360	360	360	360	360	360	360	360	360
<b>Num. groups: chain</b>	40	40	40	40	40	40	40	40	40

586  
587  
588  
589  
590  
591  
592  
593

**Table A3 Multilevel model estimates for social influence measures: (i) probability of copying the inherited solution at Trial 1, (ii) distance between the inherited solution and the solution used at Trial 1, and (iii) distance between the inherited and transmitted solutions.** Separate models are reported for each dependent variable using pooled data. All models include random intercepts for transmission chain to account for the hierarchical structure of the data. Standard errors are shown in parentheses. Statistical significance is denoted by asterisks (\*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$ ).

594 Individual Exploration

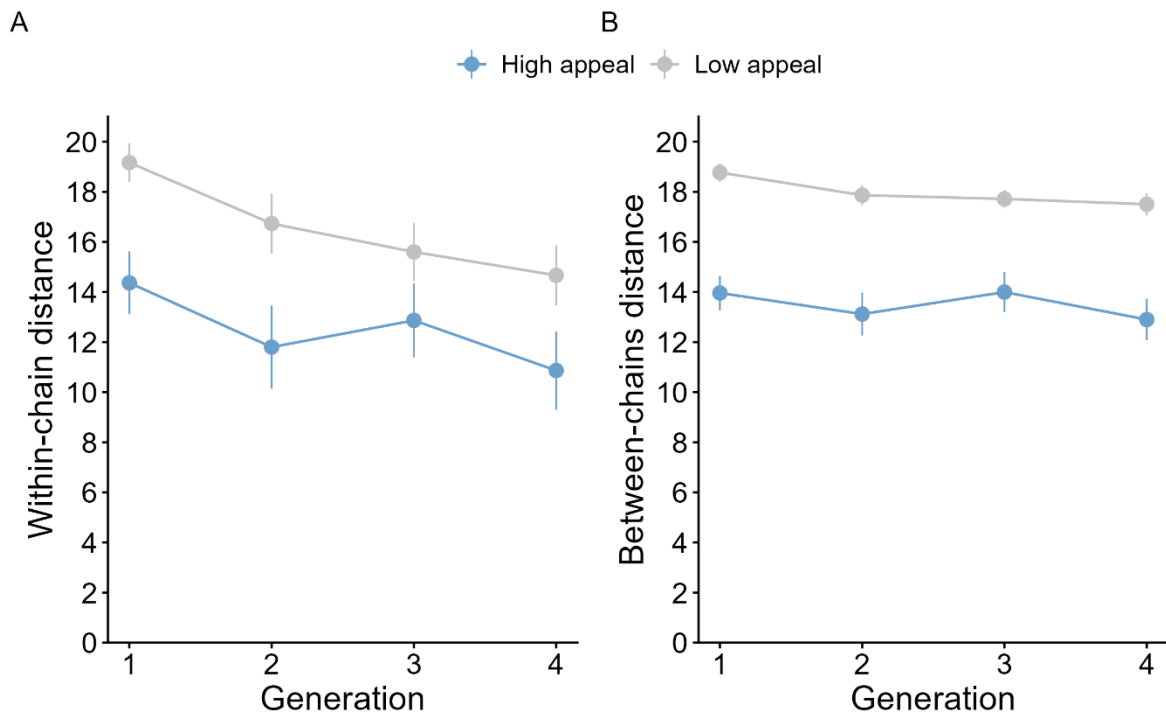


595  
596  
597 **Fig A4 Individual exploration.** (A) Mean within-individual solution distance by treatment. (B) Mean within-  
598 individual solution distance by treatment and generation. Distance is measured as the sum of absolute differences  
599 across all nutrient allocations in the solution vector (Manhattan distance) averaged over trial pairs. Error bars  
600 represent ±1 SE. Larger distances indicate greater within-individual exploration across trials.

601  
602

	Within-individual solution distance	
<b>(Intercept)</b>	12.28*** (0.83)	12.34*** (1.02)
<b>Generation</b>	-1.57*** (0.24)	-1.59*** (0.34)
<b>Low Appeal</b>	4.48*** (0.81)	4.35** (1.44)
<b>Generation x Low Appeal</b>		0.05 (0.48)
<b>Num. obs.</b>	480	480
<b>Num. groups: chain</b>	40	40

603  
604 **Table A4. Multilevel model estimates for individual exploration.** Separate models are reported for each  
605 dependent variable (Mean within-individual solution distance) using pooled data. All models include random  
606 intercepts for transmission chain to account for the hierarchical structure of the data. Standard errors are shown in  
607 parentheses, and statistical significance is denoted by asterisks (\*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$ ).



609  
 610 **Figure A5. Cultural convergence of transmitted solutions.** (A) Within-chain mean pairwise  
 611 Manhattan distance among the three transmitted solutions in each chain. (B) Between-chains mean  
 612 pairwise Manhattan distance, computed for each chain as the average distance between its three  
 613 transmitted solutions and all transmitted solutions from other chains in the same treatment-generation  
 614 group.

615  
 616  
 617  
 618  
 619  
 620  
 621  
 622  
 623  
 624  
 625  
 626  
 627  
 628  
 629  
 630  
 631  
 632  
 633

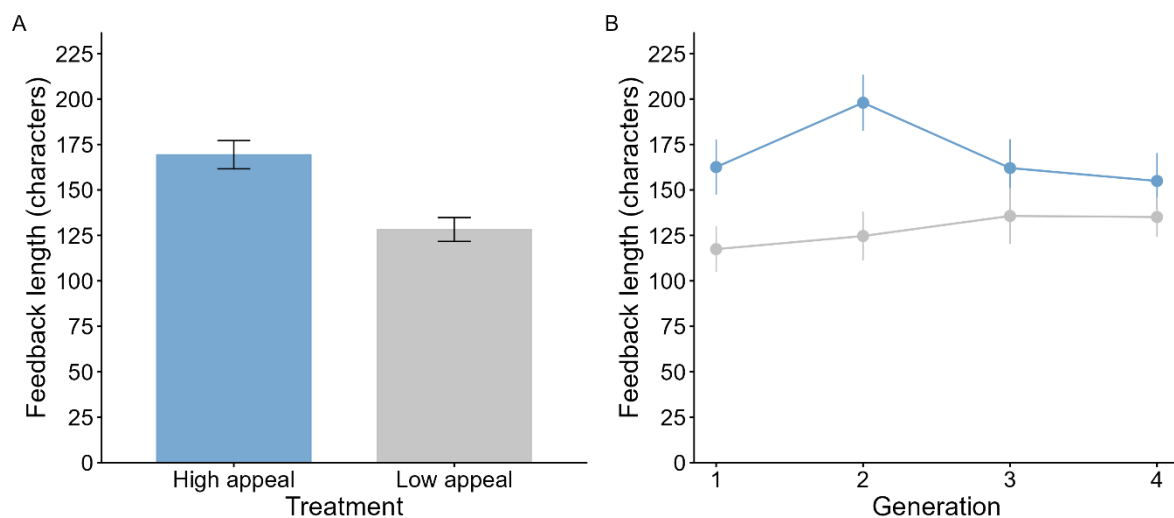
## 634 Analyses on Written Advice

### 635 Quantitative Textual Analysis

636 To complement the manual coding of advice content, we conducted quantitative analyses of  
637 the linguistic properties of participants' written advice ( $n = 480$ ; 240 per treatment). Before  
638 analysis, we removed the fixed prompt prefix ("The field yields a higher output when...") in  
639 order to reflect participants' free-response content only. We compared two text measures across  
640 treatments and generations: message length, and semantic similarity.

#### 641 *Message Length*

642 Message length was defined as the number of characters in each cleaned advice message. This  
643 measure captures the overall verbosity and degree of elaboration in communicated strategies  
644 (Figure A6).



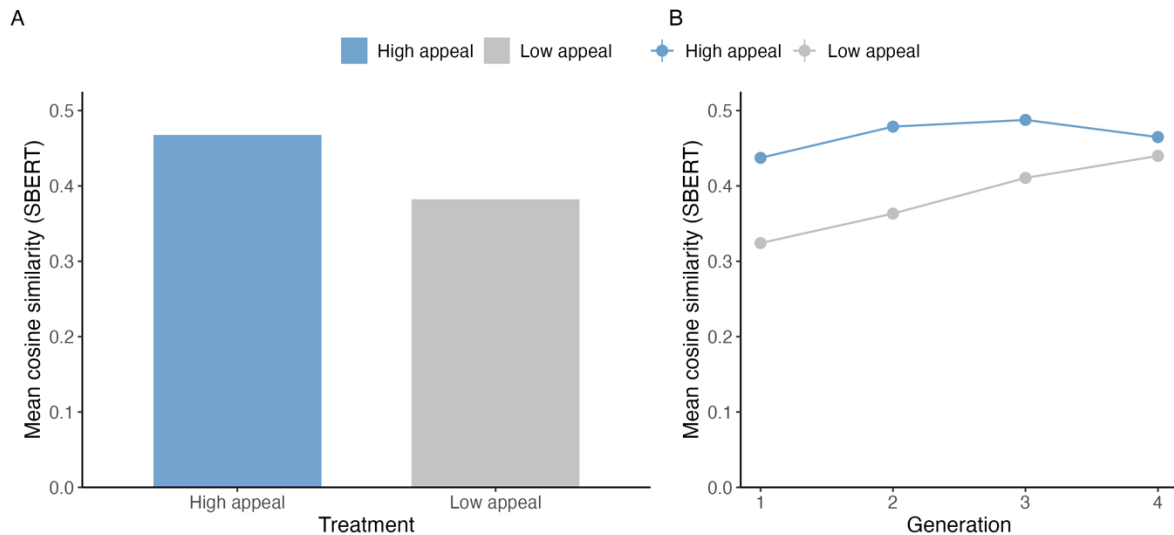
645

646 **Fig A6. Feedback length in characters generation and treatment.** (A) Mean feedback length by treatment.  
647 (B) Mean feedback length by treatment and generation. Error bars represent  $\pm 1$  SE.

#### 648 *Semantic Similarity*

649 To quantify convergence in the semantic content of advice messages, we used a sentence-  
650 embedding approach. Each advice message was transformed into a fixed-length numerical  
651 vector using a pre-trained Sentence-BERT (SBERT) model (all-MiniLM-L6-v2). We then  
652 computed pairwise cosine similarity between message vectors. Mean cosine similarity was  
653 calculated (i) within each treatment overall and (ii) within each treatment  $\times$  generation cell by  
654 averaging across all pairwise comparisons among messages. This procedure allowed us to  
655 examine whether advice became more semantically aligned over generations and whether  
656 semantic convergence differed between treatments (Figure A7).

657 Semantic similarity is initially lower in the low-appeal treatment than in the high-appeal  
658 treatment, indicating greater diversity in advice content in early generations. Across  
659 generations, similarity increases in the low-appeal condition, whereas changes in the high-  
660 appeal condition are comparatively modest. By the fourth generation, both treatments reach a  
661 comparable level of within-group semantic similarity.  
662



663  
 664 **Figure A7. Semantic similarity by treatment and generation.** Semantic similarity of transmitted feedback  
 665 messages using SBERT embeddings and cosine similarity. (A) Overall mean cosine similarity within treatment.  
 666 (B) Mean cosine similarity within each treatment x generation, shown as trajectories across generations; each  
 667 point is the mean over all message pairs within that cell. SBERT similarities are from all-MiniLM-L6-v2; higher  
 668 values indicate greater semantic overlap. Values closer to 1 indicate that two messages occupy nearby positions  
 669 in semantic space (i.e., express similar underlying ideas), whereas values closer to 0 indicate weaker semantic  
 670 overlap.

## 671 Classification of the written advice

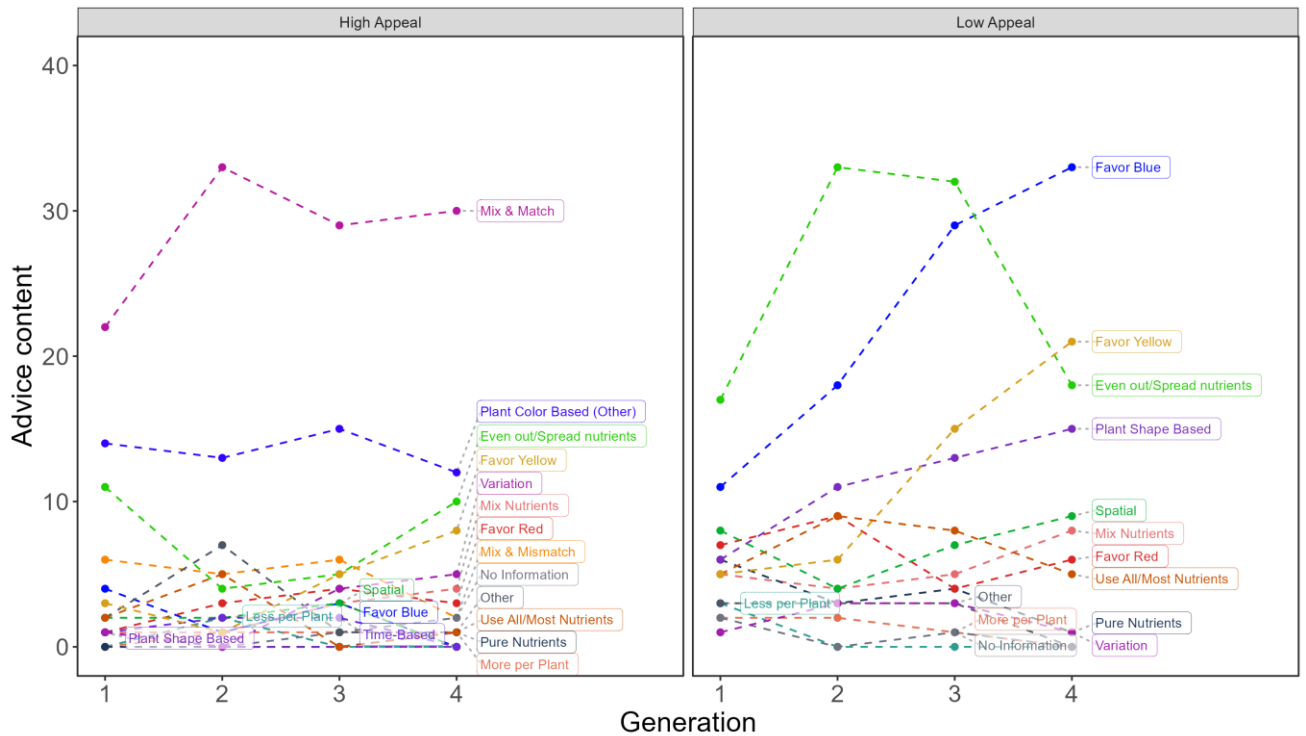
672 In addition to quantitative outcomes, we analyzed the content of participants' written advice to  
 673 classify how strategies were represented and transmitted across generations. Written advice  
 674 text pieces were coded into a set of pre-determined strategy categories. Categories that  
 675 conceptually be advised together were treated as non-mutually exclusive, allowing multiple  
 676 strategy elements to be coded within the same message when applicable. We show category  
 677 frequencies by generation and treatment condition to assess how transmitted messages evolved  
 678 over time in Figure A8. We additionally compared the manual classification to an automated  
 679 LLM-based classification using GPT-5.1. The two approaches agree with 93% accuracy, and  
 680 all results and figures are robust to using the LLM-based classification. The classifications,  
 681 along with the prompt and script used to generate the GPT-5.1 labels, are provided in the  
 682 supplementary repository.

683

Category	Description	Example Advice (The field yields a higher output when...)
<i>Mix &amp; Match</i>	Recommends matching nutrient colors to primary plant colors (blue, yellow, red) when possible and combining nutrients to recreate remaining secondary colors (orange, purple, green) based on color theory.	<i>...you match the colours according to colour theory. For example, for green, add one blue and one yellow. For purple, add one red and one blue. and so forth. For the already existing primary colours e.g. red, blue or yellow, just add one of the same colour</i>
<i>Mismatch</i>	Recommends avoiding matching nutrient colors to plant colors (blue, yellow, red), including both simple strategies (using non-corresponding colors) and more complex strategies, such as the opposite of <i>Mix &amp; Match</i> , that allocate remaining primary colors using color theory to contrast the plant's existing color.	<i>... you use the colours that do not make up the plants colour. If the plant is blue select one each for yellow and red. If the plant is orange, select one for blue as yellow + red = orange so should not be selected.</i>

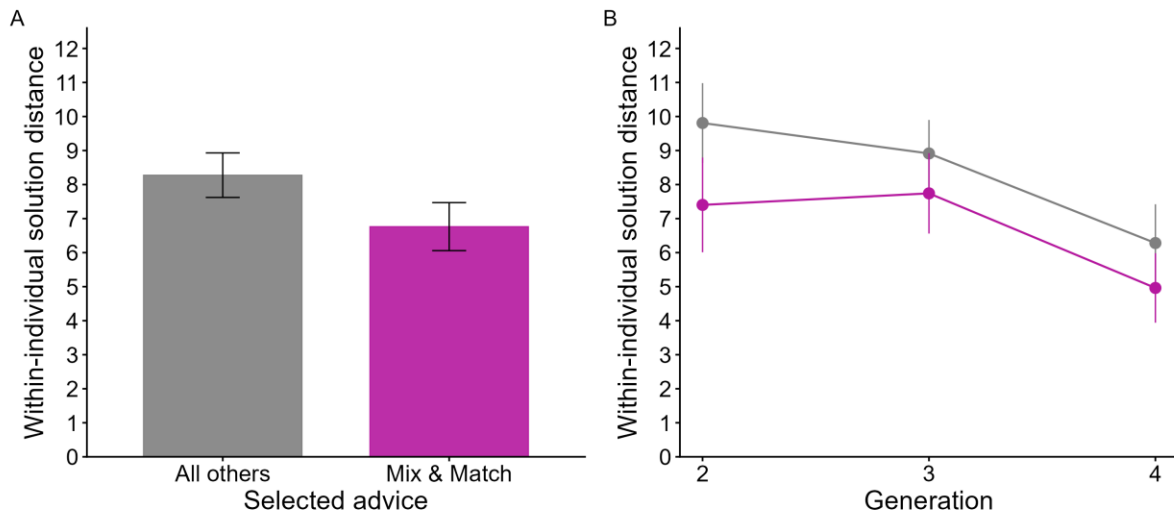
<i>(Other) Plant Color Based matching</i>	Recommends a strategy in which plant color is the primary cue for nutrient allocation (without explicit <i>Mix &amp; Match</i> or <i>Mismatch</i> strategies). Includes simple color matching for primary colors only.	<i>...you focus on matching the colours, I found you got a higher score when prioritising the colours that do not require mixing to match (e.g. giving a red plant two drops of red instead of focusing on giving orange both red and yellow)</i>
<i>Plant Shape Based Matching</i>	Recommends a strategy in which plant shape is the primary cue for nutrient allocation, including feeding plants with the same shape in the same way.	<i>...Crosses are Red, Stars are Yellow I can provide that lower output will be achieved when circles are blue and have tested triangles are blue to some succes, Hearts are red .</i>
<i>Even Out/Spread nutrients</i>	Recommends distributing nutrients broadly or evenly across plants or feeding each plant.	<i>...the nutrients are spread out.</i>
<i>Less Nutrients per Plant</i>	Recommends allocating fewer nutrients to each plant without explicitly suggesting even distribution.	<i>...you use fewer bottles of the solution.</i>
<i>More Nutrients per Plant</i>	Recommends allocating more nutrients to each plant.	<i>...you give more than one of the same nutrient per plant.</i>
<i>Mix Nutrient Colors</i>	Recommends combining multiple nutrient colors within a single plant.	<i>...you combine two different solutions on the same field</i>
<i>Pure Nutrients</i>	Recommends avoiding mixing nutrient colors within a plant.	<i>...I used individual colours by themselves rather than combining other colours together</i>
<i>Spatial Strategy</i>	Recommends space-based strategies related to the field (e.g., feeding specific rows, columns, corners, or the center in specific ways).	<i>...blue and yellow is used on the top row. Red and yellow on the middle row.</i>
<i>Use All/Most Nutrients</i>	Recommends using all or nearly all 16 available nutrients per round.	<i>...I used all the nutrients i got a higher score</i>
<i>Time-Based</i>	Recommends strategies based on speed, timing, or temporal pacing.	<i>...you move fast and submit faster</i>
<i>Favor Blue Nutrient</i>	Recommends explicitly prioritizing blue nutrients over other colors (or prioritizing blue equally with another nutrient color).	<i>...you use blue potions</i>
<i>Favor Yellow Nutrient</i>	Recommends explicitly prioritizing yellow nutrients over other colors (or prioritizing yellow equally with another nutrient color).	<i>...when I only use the yellow solution instead of only the blue or red</i>
<i>Favor Red Nutrient</i>	Recommends explicitly prioritizing red nutrients over other colors (or prioritizing red equally with another nutrient color).	<i>...red nutrients are used compared to yellow and blue nutrients</i>
<i>Variation</i>	Recommends diversifying allocations, randomizing, or changing strategies across trials.	<i>...you try to input the colors of the nutrients randomly on different shapes</i>
<i>No Information / No Advice</i>	No Information / No Advice Contains no actionable strategic information or refers only to the latest solution.	<i>...sorry, I really couldn't figure out the pattern for success</i>
<i>Other</i>	Contains advice that does not fit into the predefined categories.	<i>...you trial different combinations, find one that generally scores highly, then repeat that.</i>

684 **Table A5. Coding scheme and illustrative examples of written advice.** Advice messages were categorized into  
685 strategy types based on their substantive content. Categories are not mutually exclusive; a single message could  
686 be assigned to multiple categories where applicable. Descriptions summarize the defining features of each  
687 category, and example excerpts illustrate typical coded responses. Examples are extracted from actual participant  
688 advice keeping the original spelling and grammatical errors.  
689



690  
691  
692  
693  
694  
695  
696  
697  
698  
699

**Figure A8. Advice content across generations by treatment conditions.** Counts of advice messages coded into a predefined set of strategy categories. In the high-appeal treatment, advice promoting the costly “mix-and-match” strategy increased rapidly in the first two generations and remained relatively stable with a mild decrease. In contrast, in the low-appeal treatment, advice favoring the least costly blue nutrient increased gradually across generations. Advice recommending an even distribution of nutrients—an effective strategy in this task—was more prevalent in the low-appeal treatment.



700  
701  
702  
703  
704  
705  
706  
707

**Figure A9. Individual exploration among participants who inherited a mix-and-match solution.** (A) Mean within-individual solution distance by treatment. (B) Mean within-individual solution distance by treatment and generation. Distance is measured as the sum of absolute differences across all nutrient allocations in the solution vector (Manhattan distance), averaged over trial pairs. Error bars represent  $\pm 1$  SE. Larger distances indicate greater within-individual exploration across trials.

	<b>Within-individual solution distance</b>	
<b>(Intercept)</b>	12.77*** (1.86)	13.73*** (2.51)
<b>Generation</b>	-1.50** (0.57)	-1.81* (0.80)
<b>Selected Mix &amp; Match</b>	-1.54 (0.95)	-3.50 (3.59)
<b>Generation x Selected Mix &amp; Match</b>		0.66 (1.16)
<b>Num. obs.</b>	180	180
<b>Num. groups: chain</b>	20	20

708  
709  
710  
711  
712

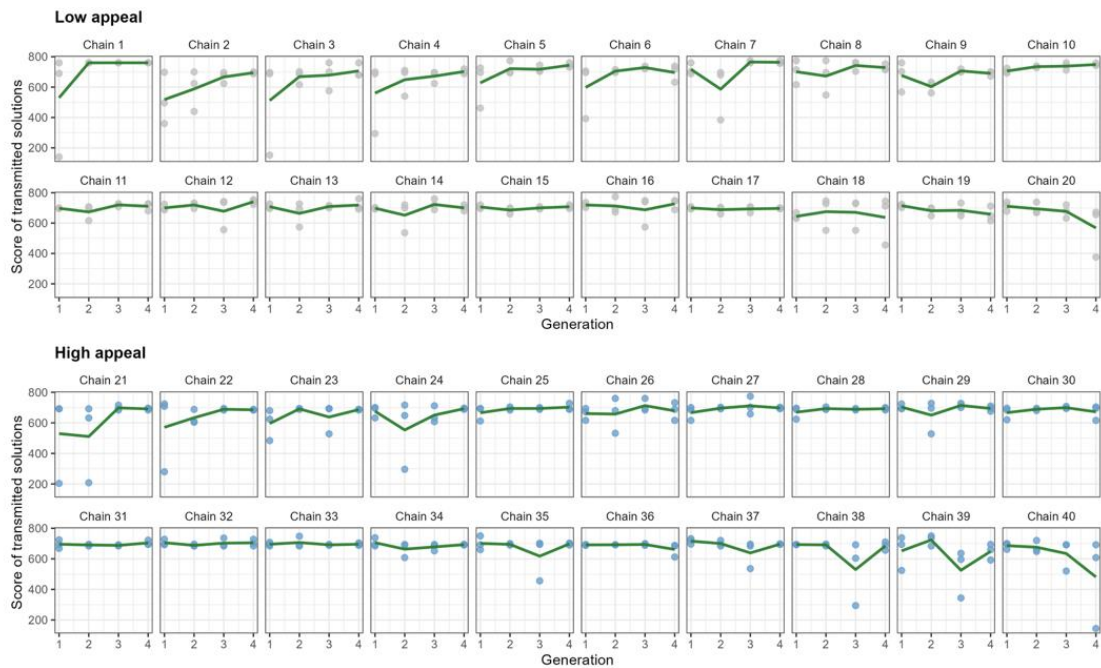
**Table A6.** Multilevel model estimates for individual exploration, controlling those who chose to inherit *Mix & Match* solution. All models include random intercepts for transmission chain to account for the hierarchical structure of the data. Standard errors are shown in parentheses and statistical significance is denoted by asterisks (\*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$ ).

713  
714  
715

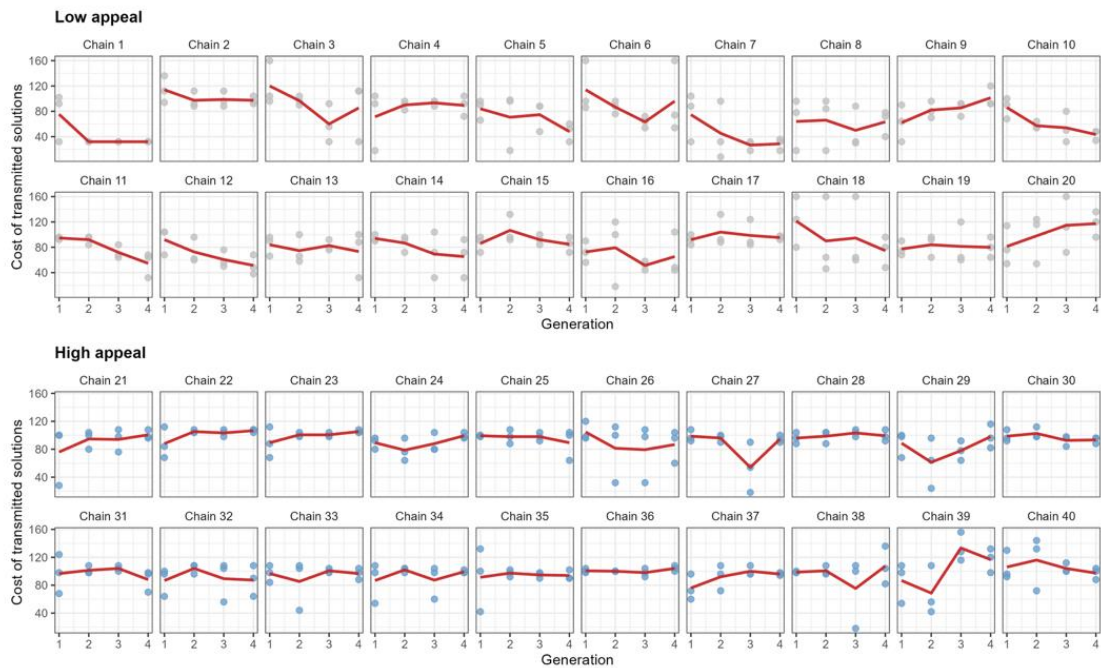
716 Scores and Costs within Transmission Chains

717

A



B



718

719

720

721

722

723

**Figure A10. Scores and cost trajectories across chains.** (A) Scores (net payoffs) of transmitted solutions within a chain. (B) Costs of transmitted solutions within a chain. Each dot represents a participant's score (or cost) of the transmitted solution. Green lines indicate the average score, and red lines indicate the average cost within the chain. Chains are ordered by their estimated slope of score across generations to facilitate comparison

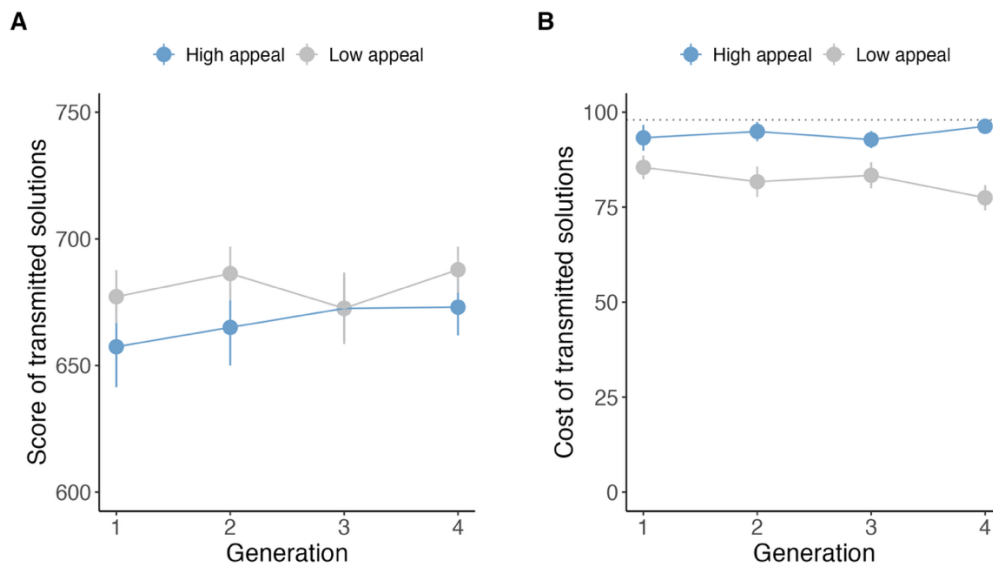
## 724 Cumulative Cultural Evolution without Payoff-Biased 725 Social Learning

726 In a previous study, we investigated a variant of this experiment that was identical to the present  
727 design except that participants were not provided with information about demonstrators'  
728 payoffs. Participants instead relied exclusively on written advice transmitted by earlier  
729 participants prior to selecting a solution. This design allowed for social learning via instruction  
730 transmission, while eliminating the possibility of payoff-biased social learning.

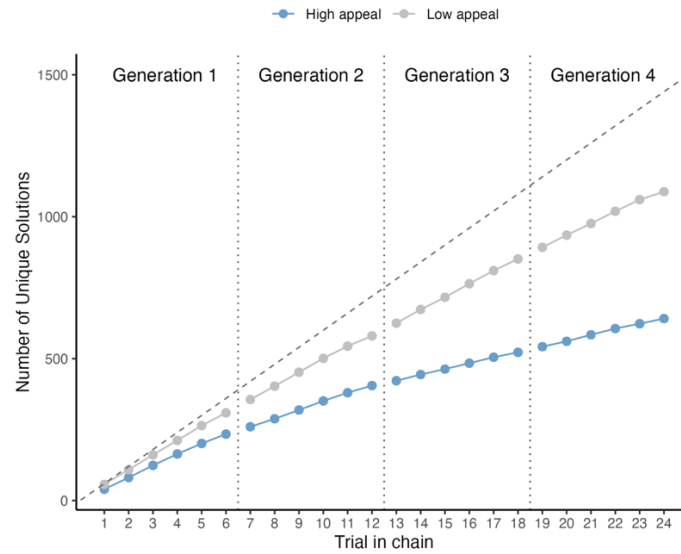
731  
732 Results show no statistically significant evidence of cumulative improvement in scores across  
733 generations in either treatment. Although the figures suggest a gradual divergence in solution  
734 costs between treatments (Figure A11), this difference is not statistically supported by the  
735 regression analyses (Table A7). We see a similar pattern of exploration in this variant of the  
736 game: In Low-Appeal treatment, participants create a higher number of unique solutions  
737 (Figure A12).

738  
739 Together, these findings suggest that in this task, payoff-biased social learning is necessary but  
740 not sufficient for cumulative cultural evolution. While payoff information can overcome payoff  
741 noise and enable adaptive refinement in environments with low cognitive appeal, strong biases  
742 in expectations of efficacy can erode its benefits, limit the exploration and allow inefficient  
743 solutions to persist even under conditions typically favorable to cumulative improvement.

744  
745



746  
747 **Figure A11. Study without demonstrator payoff information: Score and cost of culturally transmitted**  
748 **solutions across generations.** A) In the low-appeal treatment (gray line), participants showed gradual  
749 improvements in solution scores across generations, whereas scores in the high-appeal treatment (blue line) did  
750 not change significantly. B) Participants in the low-appeal treatment progressively reduced the cost of their  
751 solutions, while those in the high-appeal treatment showed no statistically significant change over time. The dotted  
752 line represents the cost of the mix-and-match solution (see Figure 3 for an illustration).



753  
754  
755  
756  
757  
758  
759

**Figure A12. Study without demonstrator payoff information: Exploration across trials.** Cumulative number of unique solutions (i.e., nutrient-to-plant allocation patterns) across trials in each treatment. The dashed black line indicates the theoretical upper bound of exploration, assuming every participant submitted a solution not previously produced in this treatment. The gray line shows the cumulative number of unique solutions observed in the low-appeal treatment, while the blue line shows the high-appeal treatment.

A) Scores			
	Low-Appeal only	High-Appeal only	Pooled
(Intercept)	676.40 *** (13.92)	653.42 *** (16.92)	653.42 *** (15.48)
Generation	1.83 (4.95)	5.44 (6.18)	5.44 (5.60)
Low Appeal			22.98 (21.90)
Generation x Low Appeal			-3.61 (7.92)
B) Costs			
(Intercept)	87.60 *** (4.60)	92.58 *** (3.22)	92.58 *** (3.97)
Generation	-2.23 (1.43)	0.70 (1.11)	0.70 (1.28)
Low Appeal			-4.98 (5.62)
Generation x Low Appeal			-2.93 (1.81)
Num. obs.	240	240	480
Num. groups: chain	20	20	40

760  
761  
762  
763  
764  
765

**Table A7. Study without demonstrator payoff information: Multilevel model estimates for expected scores (A) and costs (B) of culturally transmitted solutions across treatments.** Separate models are reported for the low-appeal condition, high-appeal condition, and pooled data to compare changes in expected scores/costs over generations, with random intercepts controlling for transmission *chain* structure of the experiment. Standard errors are shown in parentheses and statistical significance is denoted by asterisks (\*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$ ).

766

767 Experimental Instructions and Task Screenshots

768 Instructions and control questions

769

1/12

### Welcome to our study

Over the next few slides, we will explain the rules of the game and what you must do to maximize your monetary reward. Read them carefully so that you can earn as much bonus payment as possible.

[Next](#)

770

771

772

*High Appeal*

2/12

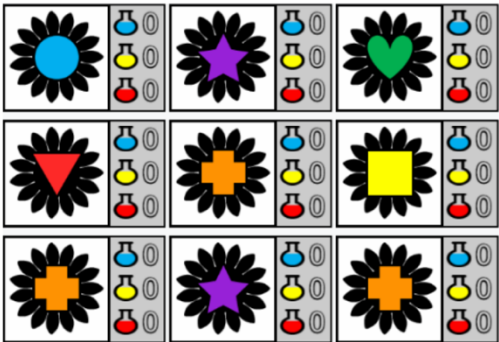
### Principle of the game

Your goal is to maximize the output of a field of 9 plants across 6 rounds.

Plants come in a variety of forms and colors and can be fed using 3 distinct nutrients.

The way you feed your plants affects the output of your field. However, we will not tell you how best to feed your plants. This is for you to figure out.

[Previous](#)



[Next](#)

773

774

775

776

*Low Appeal*

2/12

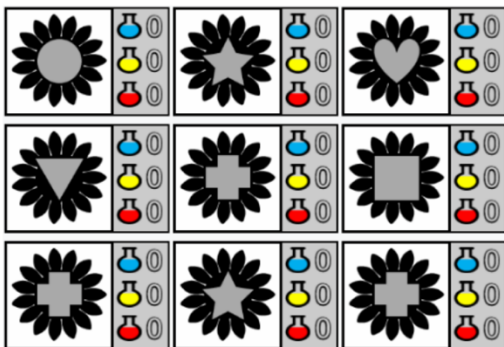
### Principle of the game

Your goal is to maximize the output of a field of 9 plants across 6 rounds.

Plants come in a variety of forms and can be fed using 3 distinct nutrients.

The way you feed your plants affects the output of your field. However, we will not tell you how best to feed your plants. This is for you to figure out.

[Previous](#)



[Next](#)

777

778

779

*High Appeal*

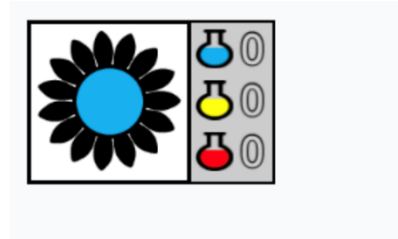
3/12

### Adding nutrients to plants

Different nutrients have different colors. In each round, you can add several droplets of nutrients to each plant. You can also mix nutrients.

Each plant has a control panel that allows you to provide droplets of nutrients to that plant. Hover your mouse over the nutrients to reveal the plant control panel.

Try it out! Feed your plant with 1 red nutrient and leave the other nutrients at 0.



Previous

You need to complete the task before continuing.

780

781

782

*Low Appeal*

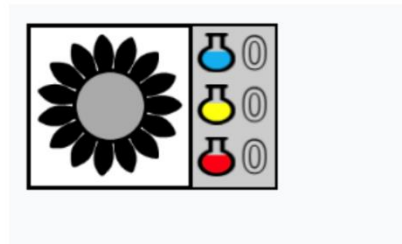
3/12

### Adding nutrients to plants

Different nutrients have different colors. In each round, you can add several droplets of nutrients to each plant. You can also mix nutrients.

Each plant has a control panel that allows you to provide droplets of nutrients to that plant. Hover your mouse over the nutrients to reveal the plant control panel.

Try it out! Feed your plant with 1 red nutrient and leave the other nutrients at 0.



Previous

You need to complete the task before continuing.

783

784

785  
786

High Appeal

4/12

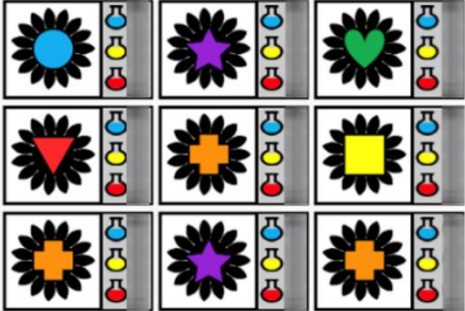
### Earning scores

At each round, you must use a total number of droplets between 1 and 16 to feed all the plants in your field. **The way you feed your plants will determine your payoff.** However, it will also depend on an element of chance. In other words, even if you choose the same solution twice, there is no guarantee that the output will be exactly the same.

Previous round

Score in the previous round  points

Your solution in the previous round:



Preview of the results after each round.  
(Actual values are obfuscated.)

Previous

Next

787  
788  
789

Low Appeal

4/12

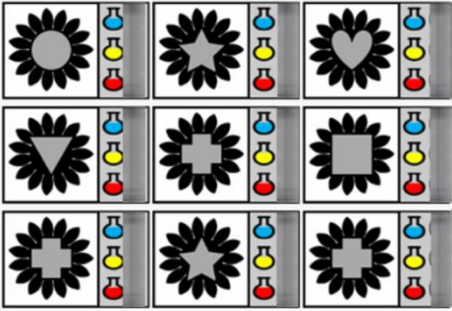
### Earning scores

At each round, you must use a total number of droplets between 1 and 16 to feed all the plants in your field. **The way you feed your plants will determine your payoff.** However, it will also depend on an element of chance. In other words, even if you choose the same solution twice, there is no guarantee that the output will be exactly the same.

Previous round

Score in the previous round  points

Your solution in the previous round:



Preview of the results after each round.  
(Actual values are obfuscated.)

Previous

Next

790  
791  
792

5/12

Please answer the questions based on the provided information. You are allowed to go back and re-read the information.

### What factors determine your score in each round?

- The way the field has been fed
- An element of chance
- Both of the above

Previous

You need to complete the task before continuing.

793

6/12

Please answer the questions based on the provided information. You are allowed to go back and re-read the information.

### How many nutrients can you use in total in each round?

- Any number
- Any number from 1 to 16
- Exactly 16

Previous

You need to complete the task before continuing.

794

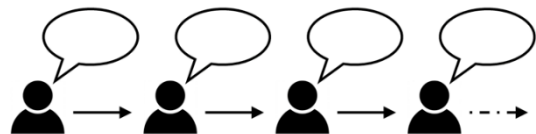
7/12

### Collective task and social information

Improving the output of the field is a collective task. Although you are playing alone, you are part of a chain of several players, each of whom also has 6 rounds to improve their solution.

You are the first player in the chain. In the last round, we will ask you to write a piece of advice for the next player and share your solution.

The number of players in a chain is not disclosed. So, there might be, or might not be, a player after you.



Previous

Next

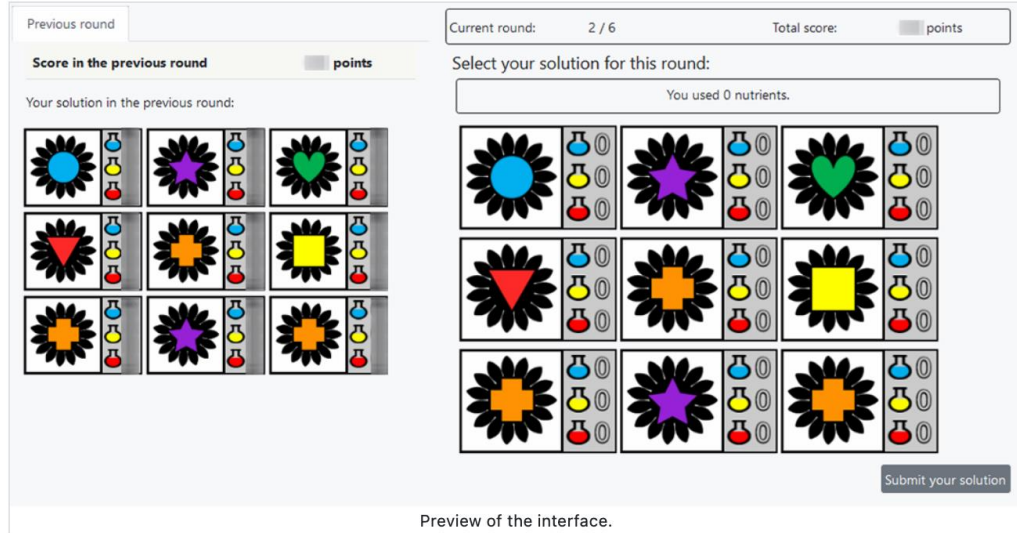
795  
796

8/12

### Interface

On the left side of your screen, a tab will provide you with information about the previous round of the game.

On the right side of your screen, you will be able to input your solution using the control panels. At any time, you can see the number of nutrients you would like to use in the current round.



Preview of the interface.

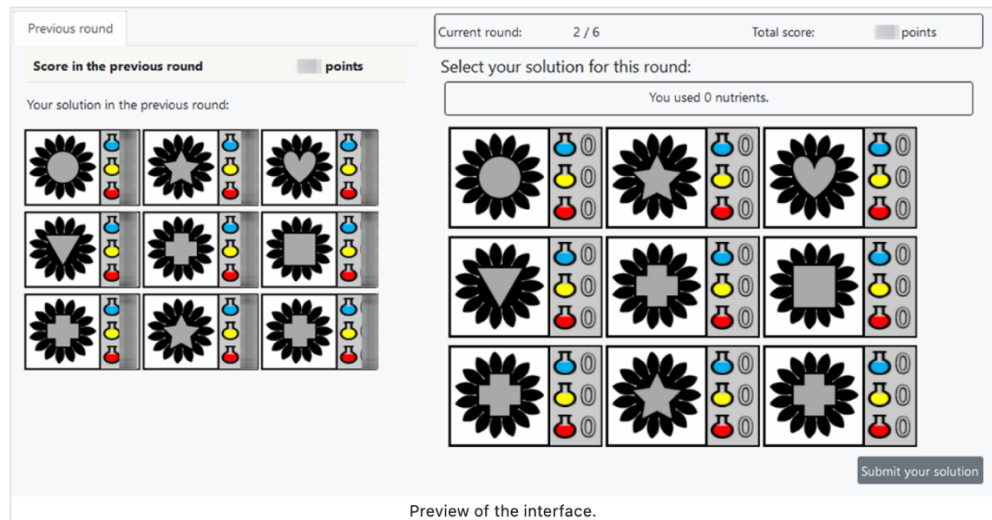
Previous

Next

### Interface

On the left side of your screen, a tab will provide you with information about the previous round of the game.

On the right side of your screen, you will be able to input your solution using the control panels. At any time, you can see the number of nutrients you would like to use in the current round.



Preview of the interface.

Previous

Next

804  
805

### Low Appeal, 2nd-4th generation

8/12

## Interface

On the left side of your screen, a tab will provide you with information about the previous round of the game. In a second tab, you will see the advice and solution from a previous player.

On the right side of your screen, you will be able to input your solution using the control panels. At any time, you can see the number of nutrients you would like to use in the current round.

Previous round (You) Advice from the previous player

Current round: 2 / 6 Total score: [ ] points

Score in the previous round [ ] points

Your solution in the previous round:

Select your solution for this round:

You used 0 nutrients.

Submit your solution

Preview of the interface.

Previous

Next

806  
807  
808

9/12

## Monetary reward

Your reward will depend on how well you did in the game and how well a future player does in their first round. We will add up the points you obtained in each round. Then, we will add the points earned by the next player in their first round to your points (unless you are the last participant in the chain).

We will then convert those points into a monetary reward in GBP. The conversion rate is:

$$2500 \text{ points} = \text{£}1.$$

This reward will be added to your participation fee on Prolific as a bonus payment.

Previous

Next

809

10/12

Please answer the questions based on the provided information. You are allowed to go back and re-read the information.

### How is your bonus payment calculated?

It will be based only on my own score.

It will be based on the score of the next player in the chain.

It will be based on my own score and the score of the next player in the first round.

Previous

You need to complete the task before continuing.

810  
811

11/12

Please answer the questions based on the provided information. You are allowed to go back and re-read the information.

### What is the equivalent of 2500 points in pounds?

£1

£1.5

£0.5

Previous

You need to complete the task before continuing.

812  
813

12/12

You have finished the instructions. You won't be able to go back to them once you start the experiment, so please make sure you have understood everything before starting and go through the instructions again if you feel you need to.

To start the experiment, click on the "Start Experiment" button below.

Previous

Start Experiment

814  
815

816 Screenshots from the gameplay

817

818

*Task Screen, Low Appeal, First generation*

Previous round

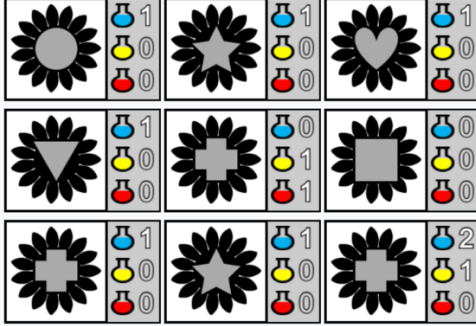
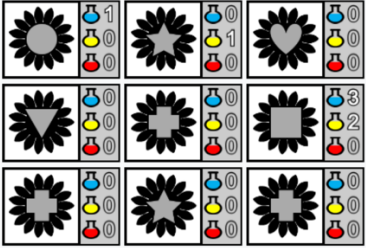
Current round: 2 / 6 Total score: 321 points

Score in the previous round 321 points

Select your solution for this round:

You used 11 nutrients.

Your solution in the previous round:



Submit your solution

819

820

821

822

*Task screen, Last round, High Appeal, First generation*

Previous round

Current round: 6 / 6 Total score: 1357 points

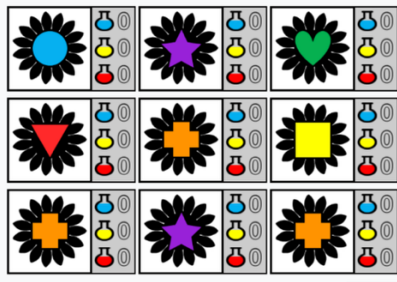
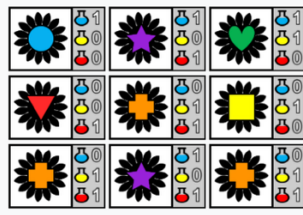
Score in the previous round 665 points

**This is your last trial.**  
This solution that you will produce at this round will be transmitted to the next participant along with a message you will write in the next step and the score you will earn from this round.  
Remember that your final monetary reward will depend on how that player performs in their first round. In order to help future participants and to maximise your own reward, it is important that your demonstration is informative.

Select your solution for this round:

You used 0 nutrients.

Your solution in the previous round:



Submit your solution

823

824

825

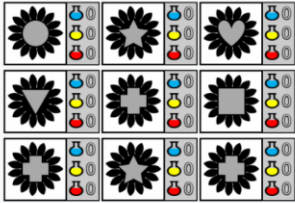
826

827  
828

### Advice Selection Screen, Low Appeal, 2nd-4th generation

Here is the field that you will have to feed. Below it are three pieces of advice from previous players who have completed 6 rounds.

Select a player to be provided with their advice and solution for the entire duration of the experiment.



Demonstration of the field.

**Player 1**  
Score at the last trial:  
706 points  
Advice written:  
*The field yields a higher output when you keep the same entry every round.*  
Select this advice

**Player 2**  
Score at the last trial:  
477 points  
Advice written:  
*The field yields a higher output when you miss at least one completely and then for a few around 1-3 have 1 at least for each colour*  
Select this advice

**Player 3**  
Score at the last trial:  
850 points  
Advice written:  
*The field yields a higher output when you use all up the droplets. Some shapes perform better with different colours. the cross performed better with yellow*  
Select this advice

829  
830  
831  
832

### Task Screen, Low Appeal, 2nd-4th generation

Previous round (You) Advice from the previous player

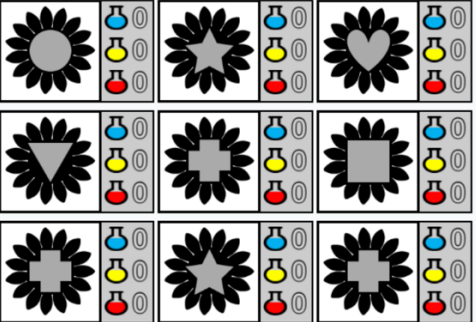
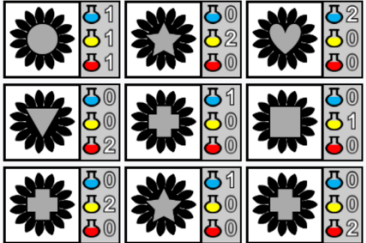
Current round: 1 / 6 Total score: 0 points

Select your solution for this round:

You used 0 nutrients.

Advice from the previous player: *The field yields a higher output when you use all up the droplets. Some shapes perform better with different colours. the cross performed better with yellow*

Solution provided by the previous player:  
Score: 850 points



Submit your solution

833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846

847  
848

### Task Screen, Low Appeal, 2nd-4th generation

Previous round (You)    Advice from the previous player    Current round: 2 / 6    Total score: 587 points

Score in the previous round: 587 points

Your solution in the previous round:


Select your solution for this round:

You used 0 nutrients.


Submit your solution

849  
850  
851

### Advice Message Screen, High Appeal

Previous round

This solution will be transmitted to the next player.  
You will see the score of this round at the end of the experiment.


Please complete the following sentence to transmit your advice to the next player. Remember that your final monetary reward will depend on how that player performs in their first round. In order to help future participants and to maximise your own reward, it is important that your advice is clear and informative.

The field yields a higher output when

Submit your advice

852  
853