

1

2

3

4

5

6

# 7 Simulating cultural and biological evolution through 8 large-scale online experiments

9

10 Thomas J. H. Morgan<sup>1+\*</sup>, Jordan W. Suchow<sup>1+\*</sup> & Thomas L. Griffiths<sup>1\*</sup>

11

<sup>12</sup> <sup>1</sup> UC Berkeley, Department of Psychology, Tolman Hall, Berkeley, 94720

<sup>13</sup> + denotes joint first authorship

<sup>14</sup> \* correspondence: [thomas.morgan@berkeley.edu](mailto:thomas.morgan@berkeley.edu), [suchow@berkeley.edu](mailto:suchow@berkeley.edu)

15

16 v1.0

17

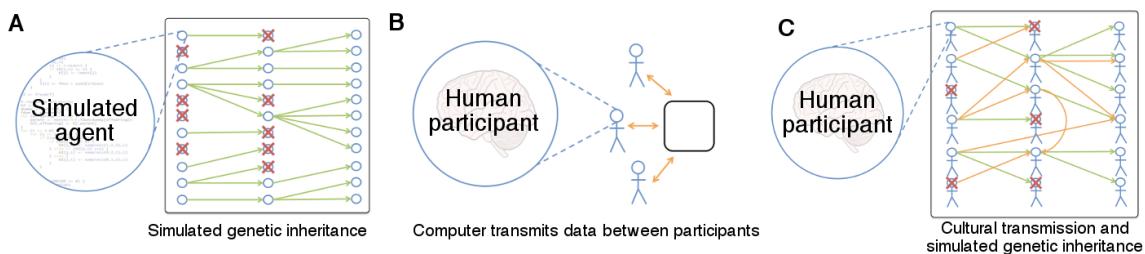
18    ***Abstract***

19    **Human cognition and behavior result from complex interactions between cultural  
20    and biological evolution, obscuring their origins. Existing approaches to solving this  
21    problem use theoretical models to generate hypotheses that are then tested in  
22    laboratory experiments. However, theoretical models make strong simplifying  
23    assumptions about the nature of human minds in order to be tractable, and  
24    laboratory experiments do not reveal evolutionary dynamics. Here, we overcome  
25    these limitations by inserting human participants into large-scale evolutionary  
26    simulations, assigning them artificial genes that modify the tasks they are asked to  
27    perform. In this way, cognitive capacities can be masked, enhanced, and transformed  
28    as if through biological evolution. We use this method to replicate and extend  
29    theoretical results concerning the evolution of social learning, the impact of learning  
30    on genetic evolution, and the co-evolutionary dynamics of learning and memory.**

31    Preserved remains provide only limited information about the evolution of human  
32    behavior and cognition. As a result, such questions are typically addressed indirectly by  
33    combining theoretical evolutionary models with laboratory experiments. The former  
34    generate predictions and identify candidate behaviors; the latter compare human behavior  
35    with these predictions. For example, theoretical models of social learning strategies have  
36    received extensive empirical verification (1–3). However, each approach has its limitations.  
37    Although theoretical evolutionary models can expose the conditions under which a  
38    candidate trait will evolve, their creation requires simplifying assumptions about the  
39    behavior or cognitive ability in question. Thus, although they illustrate the evolutionary

40 consequences of an assumed trait, their ecological validity in the specific context of human  
41 evolution remains unclear. Empirical studies, unlike theoretical models, examine the  
42 behavior of organisms directly and thus require fewer assumptions about their behavior.  
43 But they suffer from a different limitation: they provide only a static snapshot of human  
44 behavior or cognition, with no means to study evolutionary dynamics.

45 Here, we combine the strengths of theoretical models and empirical studies through  
46 large-scale online evolutionary simulations. Our technique assigns thousands of  
47 participants artificial genes that influence the structure of the task they are asked to  
48 perform, inserts them into an evolving population, and then uses their behavior to  
49 determine their fitness and hence simulated reproductive success (



50  
51 Figure 1). Use of human participants rather than mathematical abstractions or computer-  
52 based agents is advantageous because we need make no simplifying assumptions about the  
53 nature of the cognitive processes involved. Because modern humans already possess the  
54 traits we seek to explain, it is possible to construct a mapping from artificial genes to  
55 experimental tasks that masks participants' abilities (e.g., using desaturated images to  
56 mask color vision) and then explore the circumstances under which selection unmasks  
57 those abilities.

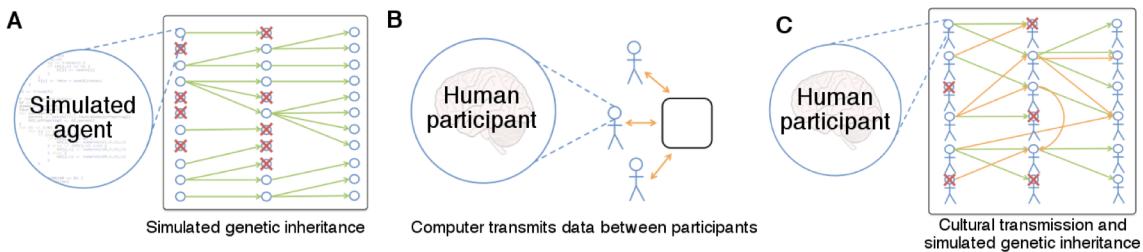


Figure 1. Three approaches to studying the evolution of human cognition. **(A)** Theoretical evolutionary models such as agent-based modeling simulate artificial agents embedded in an environment, making strong assumptions about their psychology. **(B)** Laboratory experiments examine the behavior of individual human participants. **(C)** Our approach embeds human participants, endowed with artificial genes, into large-scale online simulations of evolutionary dynamics orchestrated by a computer.

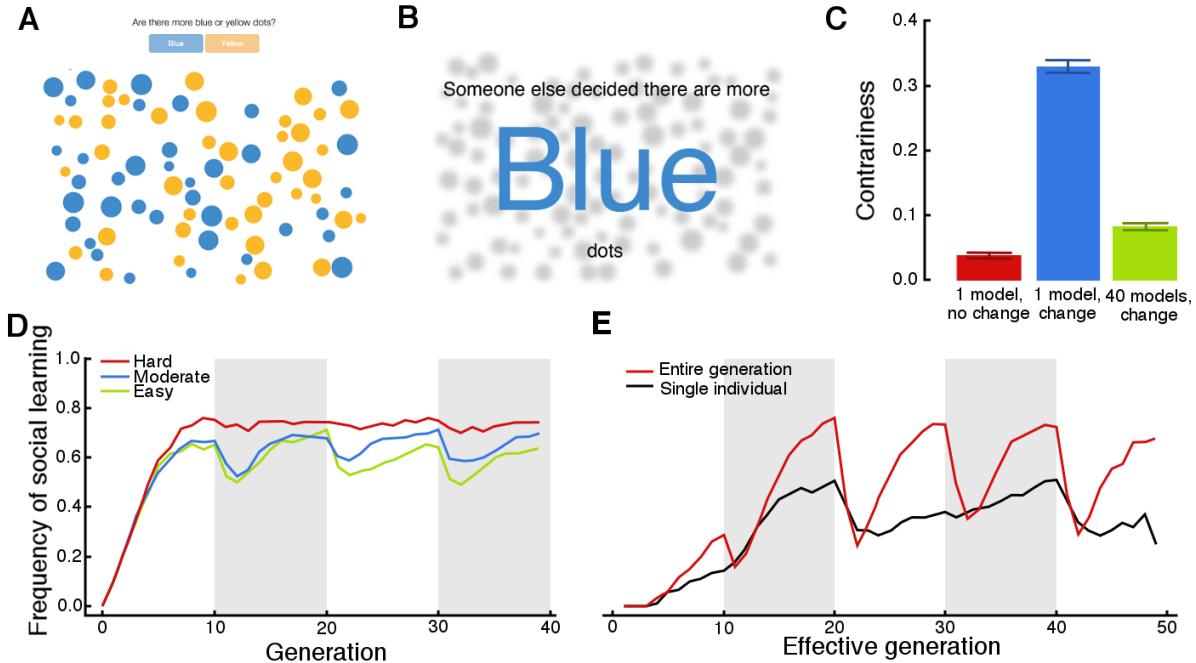
58

59        In what follows, we use this method to replicate and extend theoretical results  
 60      concerning the evolution of social learning in a changing environment, the impact of  
 61      learning on genetic evolution, and the co-evolutionary dynamics of learning and memory.  
 62      In each case, we demonstrate that theoretical results hold in evolutionary simulations with  
 63      human agents and that nuances of human behavior lead to phenomena unanticipated by  
 64      theory.

65        Although humans are adept inventors, we are perhaps best defined by our ability to  
 66      learn from others (4). Social learning enables innovations to accumulate in a population,  
 67      leading to technologies that go beyond what any one person could create. Although culture  
 68      might appear to be inherently advantageous, early theoretical work found that as the  
 69      frequency of social learning increases in a population its fitness declines until, at  
 70      equilibrium, average fitness is no better than in a population of asocial learners (5). This  
 71      result, known as *Rogers' Paradox*, arises because social learning recycles existing  
 72      information and thus cannot track environmental change.

73        To verify that our approach reproduces well-established evolutionary dynamics, we

74 first sought to recreate Rogers' Paradox. In the experiment, 1600 participants took part in  
75 125 parallel simulations, each involving 40 generations of 40 agents. Within each  
76 simulation, participants played the role of a single agent and, according to an artificial gene  
77 assigned to them, were either an asocial or a social learner. Asocial learners viewed an  
78 array of blue and yellow dots for 1s and decided which color was more numerous (**Error!**  
79 **Reference source not found.A**). Social learners made the same decision, but did not see  
80 the dots directly, instead observing the decision of a single agent selected randomly from  
81 the previous generation (**Error! Reference source not found.B**). The difficulty of the task  
82 varied across simulations (3 levels: easy, moderate, and hard, see SI). Every 10 generations,  
83 and unbeknownst to participants, the environment changed state: all the dots switched  
84 colors. Accordingly, the information acquired by social learners could be outdated. The  
85 fitness of each agent was a function of whether their decision matched the current state of  
86 the environment, with asocial learners paying a cost to observe the current state. Agents  
87 inherited their artificial gene from an agent in the previous generation, chosen with  
88 probability proportional to fitness. Inheritance was subject to mutation with a 10% chance  
89 that an agent's strategy would differ from that of their parent. The first generation  
90 consisted only of asocial learners.



91

92 Figure 2. The evolution of social learning in a changing environment. (A) The perceptual task faced by  
 93 human participants when agents learn asocially. On some trials, participants judged whether a display  
 94 contained more blue or yellow dots. (B) Social information. On some trials, participants were told the  
 95 decision of an individual from the previous generation. (C) Contrariness. Knowledge of the possibility of  
 96 environmental change increased the probability that a social learner adopted the opposite decision of  
 97 their demonstrator: without knowledge: 3.7% [3.3%, 4.2%], with knowledge: 33% [32%, 34%]. Despite the  
 98 genuine possibility of environmental change, this decreased the frequency of social learning. When social  
 99 learners could observe the entire previous generation, contrariness (i.e., adopting the minority opinion)  
 100 was less prevalent: 8.3% [7.8%, 8.8%]. (D) During periods of environmental stability (within white or grey  
 101 regions), the frequency of social learning increased, but following environmental change (the border  
 102 between white and grey regions) it decreased. The relative resilience of social learning when the learning  
 103 problem was hard is due to the low performance of the asocial learners (asocial performance: easy: 98%  
 104 [98%, 98%], moderate: 84% [84%, 85%], hard: 66% [65%, 67%]), which reduces the fitness of asocial  
 105 learners and, by increasing the amount of inaccurate information in the population prior to change,  
 106 increases the accuracy of the information after environmental change relative to easier conditions. (E)  
 107 Access to the decisions of the entire previous generation increased the peak frequency of social learning.  
 108 However, it also increased the impact of environmental change.

109 Replicating previous theoretical results, we found that social learning increased  
 110 during periods of stability, and decreased following environmental change (Figure 2D). The  
 111 average frequency of social learners after 10 generations of stability was 70% [69%, 72%]  
 112 (median and 95% credible interval), dropping to 61% [60%, 63%] two generations after  
 113 environmental change. The difficulty of the learning problem affects the evolutionary  
 114 dynamics: harder learning problems led to higher average frequencies of social learning

115 (hard: 74% [72%, 76%], moderate: 65% [62%, 67%], easy: 60% [58%, 62%], hard-  
116 moderate: 9% [7%, 12%], moderate–easy: 5% [2%, 7%]) and smaller drops in social  
117 learning in response to environmental change (easy: 16% [14%, 19%], moderate: 12%  
118 [10%, 14%], hard: 3% [1%, 5%]).

119 Social learning strategies (1) such as copying the decisions of successful individuals  
120 or disproportionately adopting the majority decision (“conformist transmission” (3)), have  
121 been identified as possible solutions to Rogers’ Paradox because they increase the efficacy  
122 of social learning. However, these strategies can also be deleterious. Conformist  
123 transmission, for example, is particularly harmful after environmental change because it  
124 prevents the spread of new ideas. Humans engage in conformist transmission (2), but there  
125 is some evidence they adjust their social learning to account for environmental change (6).  
126 Accordingly, we explored the effect of environmental change on the evolutionary dynamics  
127 of social learning when the social context is enriched by informing participants of the  
128 decisions of multiple other individuals, thus allowing more complex responses to  
129 consensus, such as conformist transmission. We recruited 1600 participants to take part in  
130 a further 89 repeat simulations. All simulations were set to moderate difficulty and  
131 participants were informed of the environmental change. Across repeats, we varied  
132 whether social learners saw the decisions of the entire previous generation (40  
133 individuals) or of a single demonstrator selected at random, as in the previous experiment.  
134 If people can tailor their social learning to environmental change, it will have less impact on  
135 the frequency of social learners.

136 We found that, although access to the entire previous generation's decisions  
137 increased the frequency of social learning when the environment was stable, it exacerbated  
138 the drop in social learning that follows environmental change (single demonstrator, prior  
139 to change: 47%, [44%, 49%], after change: 34%, [32%, 36%], drop: 13%, [11%, 15%], 40  
140 demonstrators, prior to change: 74%, [72%, 76%], after change: 28%, [26%, 30%], drop:  
141 46%, [44%, 47%] Figue 2E). This is because when participants could observe the entire  
142 previous generation their behavior was consistent with conformist transmission. Despite  
143 this, there is some evidence that participants adjusted their behavior to the risk of  
144 environmental change. For instance, when shown only a single demonstrator, participants  
145 were more likely to be contrary when aware of the risk of environmental change than  
146 when they were not aware (Figure 2C). However, the level of contrariness when the entire  
147 previous generation was available was consistently low (Figure 2C). It seems the  
148 magnitude of the social information overpowered the risk of environmental change.

149 At the end of the 19<sup>th</sup> century, James Baldwin, along with several others, proposed a  
150 mechanism by which learned behaviors interact with genetic evolution (7, 8). Originally  
151 known as organic selection, and now more widely known as the *Baldwin Effect*, it is a  
152 process by which learned behaviors are increasingly reliably acquired because of  
153 accumulated genetic change that favors their acquisition. The Baldwin Effect has been  
154 proposed as an important factor in human evolution, particularly in the evolution of  
155 language (9, 10). However, it has yet to be demonstrated in a human population.

156 To address this, we carried out an experiment in which 2400 participants each took  
157 part in 18 parallel simulations, each involving 40 generations of 60 agents. Within each  
158 simulation, participants completed a category-learning task, categorizing 8 objects that

159 varied along three binary dimensions into one of two categories (based on (21)). Following  
160 the theoretical model in (11), agents were assigned an artificial genome with 8 genes, each  
161 of which corresponded to one of the objects and had two alleles: a neutral allele that was  
162 inert, and a beneficial allele that corrected categorization errors. Fitness was determined  
163 by an agent's accuracy in categorizing the objects. Agents reproduced sexually, with an  
164 agent's two parents chosen from the previous generation with probability proportional to  
165 fitness and with each gene being inherited from a randomly selected parent. Each gene had  
166 a 10% chance of mutation, which produced a beneficial allele 25% of the time.

167 Consistent with theoretical analyses of the Baldwin Effect, the equilibrium was one  
168 in which agents had only a subset of beneficial alleles (average beneficial allele frequency  
169 after 40 generations: 46% [45%, 48%]). This is because human learning abilities make it  
170 possible to perform well with only moderate levels of genetic assistance.

171 Not all regularities are equally easy to learn, which enables us to explore the effect  
172 of learnability on the Baldwin Effect. Following (12), we considered three rule types  
173 (**Error! Reference source not found.A-C**): Type I rules were easiest to learn (99% correct  
174 [99%, 99%] unaided by genes), Type II rules were slightly harder (89% [88%, 89%]), and  
175 Type V rules were harder still (85% [84%, 85%]). We found that the beneficial allele was  
176 more frequent when the learning problem was more difficult: highest with a Type V rule  
177 (60% [59%, 63%]), lower with Type II (0.42% [0.40%, 0.44%]), and lower still with Type I  
178 (28% [25%, 30%];**Error! Reference source not found.D**).

179

180 The ability to access human learning directly allows us to answer another question:  
181 What will the Baldwin effect affect? For example, if the human capacity for language were

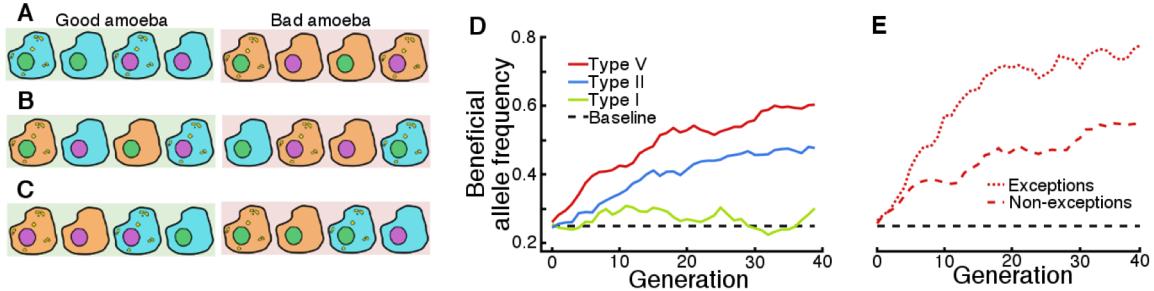


Figure 3. Experimental evidence of the Baldwin Effect and its dependence on the difficult of learning. (A) A Type I rule, in which category membership depends only on a single dimension: “blue amoeba are good and green amoeba are bad.” (B) A Type II rule, in which category membership depends on two dimensions: “orange amoeba with green nuclei and blue amoeba with purple nuclei are good”. (C) A Type V rule, in which category membership depends on all three dimensions. Type V rules superficially resemble a Type I rule, but include two exceptions to the rule. In this case, the rule is “amoeba with purple nuclei are good, unless they are blue and don’t have spots, in that case the amoeba with a green nucleus is good”. (D) The Baldwin Effect depends on the difficulty of learning, with a greater genetic response in the context of harder learning problems. (E) . In the case of Type V rules, the beneficial allele was most prevalent at loci corresponding to exceptional amoeba (76% [73%, 79%] v. 55% [53%, 57%], difference = 21% [18%, 25%]). Again, this was driven by a difference in the learning ability of participants: accuracy for non-exception amoeba was 88% [87%, 88%], while for exception stimuli it was 73%, [72%, 74%] (difference = 15% [14%, 15%])

182 to have resulted from such a process, what form of genetic influence might we expect on  
183 the development of language? To this end, we tested whether the genetic response can be  
184 tailored to the more difficult components of a task. In particular, note that Type V rules can  
185 be described as a simple rule with a pair of exceptions; these exceptions are particularly  
186 hard to learn (non-exception accuracy: 88% [87%, 88%], exception accuracy: 73%, [72%,  
187 74%]). Thus the difficulty of the task varied across the eight objects. We found that the  
188 beneficial allele reached higher frequency at loci corresponding to exception objects (76%  
189 [73%, 79%]) than at the other loci (55% [53%, 57%]), **Error! Reference source not**  
190 **found.**E). Thus, extrapolating to the case of language, we might expect features that were  
191 most problematic for human learning mechanisms to be supported genetically.

192 There is a broad consensus that the human mind represents partially independent  
 193 faculties that co-evolved (4, 13), with the benefits of each supporting the others. Consider,  
 194 for example, the coevolution of memory and learning in the context of sequential decision-  
 195 making problems (14). Because greater investment in learning favored extending the  
 196 temporal capacity to store information (i.e., memory), which in turn favored further  
 197 investment in learning, memory and learning co-evolved. Environmental change disrupted  
 198 this co-evolution by reducing the utility of memory. However, the more aspects of cognition  
 199 that are considered, the more assumptions are required, meaning that such coevolutionary  
 200 models represent a particular challenge for theoretical work. As such, using human  
 201 participants in this context is particularly valuable. Moreover, more realistic versions of  
 202 learning and memory may allow the co-evolution to occur despite a changing environment.  
 203 Accordingly, we tested whether these results hold true for human behavior.

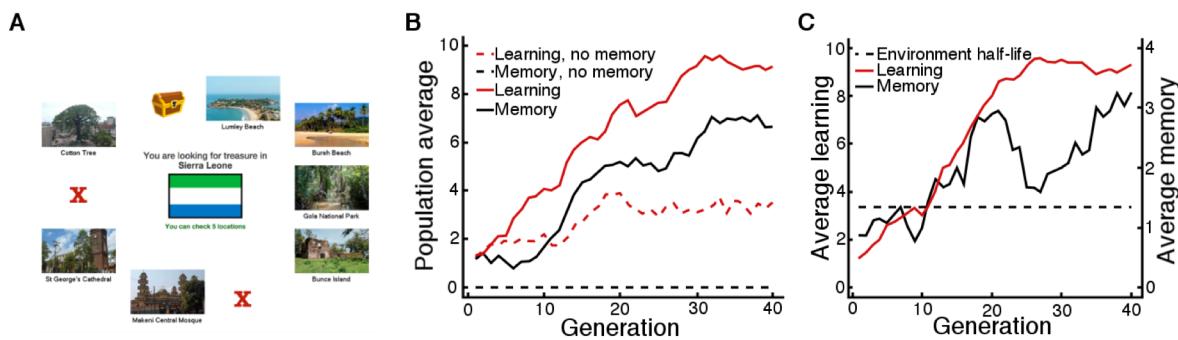


Figure 4. Experimental co-evolution of learning and memory. (A) The experimental task. On each trial, participants visited a country (in the case shown, Sierra Leone) and searched for treasure hidden in one of 10 locations. Participants could check a number of locations before making a decision, as specified by their learning gene. Checking a location revealed whether the treasure was there (as indicated by the treasure chest) or not (the red "X"s). After checking locations, the participants made their final decision for that round. (B) When both learning and memory were permitted (solid lines), they co-evolved, with learning reaching a greater equilibrium value with memory than it did without it. (C) Reducing the constraints on behavior enabled learning and memory to co-evolve even when the environment was unstable.

204

205 We carried out three experiments, each with 1600 participants, simulating an

206 evolutionary process with 40 generations of 40 agents. Within each simulation,  
207 participants played 20 trials of a sequential decision-making task in which they searched  
208 for treasure at different locations within a country. There were 4 different countries, each  
209 with 10 locations, only one of which produced a reward. Agents were assigned two genes: a  
210 learning gene that controlled the number of locations that could be tested before making a  
211 selection, and a memory gene that controlled the elapsed time after which a previously  
212 visited country would become unrecognizable (it would change to a new country). Fitness  
213 was a function of the accumulated reward and the costs associated with memory and  
214 learning. Agents inherited their genes from a parent in the previous generation chosen with  
215 probability proportional to fitness. Each gene mutated with probability 50%, causing its  
216 value to increase or decrease by 1.

217 In the first two experiments, we verified that learning and memory co-evolve in a  
218 static environment when there are heavy constraints on participants' behavior, matching  
219 those of the theory. Specifically, participants checked the number of locations specified by  
220 their learning gene and, when revisiting a familiar country, relied entirely on their memory.  
221 In one of the two simulations, we fixed memory to zero; in the other, it was allowed to  
222 evolve. Matching the theoretical model, we found that when both learning and memory  
223 were permitted, they evolved to high equilibria [9.15, [8.74, 9.55] locations and 6.67 [6.21,  
224 7.14] trials, respectively]. Without memory, learning evolved to a lower equilibrium value  
225 of 3.60 [3.18, 4.03] locations (Figure 4B), providing strong evidence that learning and  
226 memory co-evolved.

227 In the third experiment, we examined whether the co-evolution could proceed in the  
228 face of environmental change when more complex behaviors are permitted. Participants

229 could now check any number of locations, up to a maximum determined by the value of  
230 their learning gene; this was true both of new and familiar countries. We set the probability  
231 of environmental change to 40%, such that at the end of every trial, there was a 40%  
232 chance that the treasure would move. Modeling results (see SI) showed that, if participants'  
233 behavior were consistent with the constraints in the previous experiments, this rate of  
234 environmental change prevents evolution of memory or learning. In contrast, we found  
235 that the co-evolution proceeded in the face of environmental change when real human  
236 participants were involved, with the final value of memory being around 2.5 times the  
237 environmental half-life (final generation, learning: 9.30,[8.97, 9.63] locations, memory:  
238 3.25 [2.77, 3.74] trials, Figure 4C).

239 We have shown that the insertion of human participants into large-scale  
240 evolutionary simulations facilitates the experimental study of the genetic and cultural  
241 evolution of cognition and behavior. Constraining human behavior to match the  
242 assumptions of theoretical models reproduces theoretical results such as Rogers' Paradox,  
243 the Baldwin Effect, and the co-evolution of learning and memory. Relaxing these  
244 constraints leads to new insights. Although people adjust their social learning in light of  
245 environmental change, they do so ineffectively, lessening the frequency of social learning.  
246 The Baldwin effect is more likely to genetically support behaviors that are harder to learn.  
247 Finally, human behavior enables learning and memory to coevolve despite rapid  
248 environmental change. Collectively, these results support large-scale evolutionary  
249 simulations as a paradigm for investigating human cultural and biological evolution,  
250 complementing approaches based on theoretical models and laboratory experiments.

251

252 References

- 253 1. L. Rendell *et al.*, Cognitive culture: theoretical and empirical insights into social  
learning strategies. *Trends Cogn. Sci.* **15**, 68–76 (2011).
- 254 2. T. J. H. Morgan, L. E. Rendell, M. Ehn, W. J. E. Hoppitt, K. N. Laland, The evolutionary  
basis of human social learning. *Proc. R. Soc. B Biol. Sci.* **279**, 653–662 (2011).
- 255 3. R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (University of Chicago  
Press, Chicago, 1985).
- 256 4. R. Boyd, P. J. Richerson, J. Henrich, The cultural niche: why social learning is essential  
for human adaptation. *Proc. Natl. Acad. Sci.* **108**, 10918–10925 (2011).
- 257 5. A. R. Rogers, Does Biology Constrain Culture. *Am. Anthropol.* **90**, 819–831 (1988).
- 258 6. U. Toelch, M. J. Bruce, M. T. H. Meeus, S. M. Reader, Humans copy rapidly increasing  
choices in a multiarmed bandit problem. *Evol. Hum. Behav.* **31**, 326–333 (2010).
- 259 7. B. Weber, D. Depew, *Evolution and Learning: The Baldwin effect Reconsidered* (MIT  
press, 2003).
- 260 8. J. Baldwin, A New Factor in Evolution. *Am. Nat.* **30**, 441–451 (1896).
- 261 9. S. Pinker, P. Bloom, natural language and natural selection. *Behav. Brain Sci.* **13**, 707–  
784 (1990).
- 262 10. T. W. Deacon, *The Symbolic Species: The co-evolution of language and the brain* (W. W.  
Norton, New York, NY, 1997).
- 263 11. G. Hinton, S. Nowlan, How learning can guide evolution. *Complex Syst.* **1**, 495–502  
(1987).
- 264 12. R. N. Shepard, C. I. Hovland, H. M. Jenkins, Learning and Memorization of

- 274            Classifications. *Psychol. Monogr. Gen. Appl.* **75** (1961),  
275            doi:10.1017/CBO9781107415324.004.
- 276        13. S. Pinker, The cognitive niche: coevolution of intelligence, sociality, and language.  
277            *Proc. Natl. Acad. Sci.* **107 Suppl**, 8993–9 (2010).
- 278        14. B. Kerr, M. W. Feldman, Carving the cognitive niche: optimal learning strategies in  
279            homogeneous and heterogeneous environments. *J. Theor. Biol.* **220**, 169–188 (2003).
- 280
- 281        Acknowledgments
- 282        We thank J. Hamrick, M. Pacer, S. Meylan, C. Ewing and A. Mitchell for their assistance in  
283            designing the experimental software. This work was funded by NSF grants 1456709 and  
284            1408652.

1

2

3

4

5                   Simulating cultural and biological evolution  
6                   through large-scale online experiments

7                   ~ Supplementary Information ~

8

9

10

11                  Thomas J. H. Morgan<sup>1+\*</sup>, Jordan W. Suchow<sup>1+\*</sup> & Thomas L. Griffiths<sup>1\*</sup>

12

13                  <sup>1</sup> UC Berkeley, Department of Psychology, Tolman Hall, Berkeley, 94720

14                  + denotes joint first authorship

15                  \* correspondence: [thomas.morgan@berkeley.edu](mailto:thomas.morgan@berkeley.edu), [suchow@berkeley.edu](mailto:suchow@berkeley.edu),

16                  [tom\\_griffiths@berkeley.edu](mailto:tom_griffiths@berkeley.edu)

17

18

19	<b>1 Full methods</b>	<b>3</b>
20	<b>1.1 The evolution of social learning</b>	<b>3</b>
21	<b>1.1.1 Experiment 1</b>	<b>3</b>
22	<b>1.1.1.1 The experimental task</b>	<b>3</b>
23	<b>1.1.1.2 The experimental procedure</b>	<b>5</b>
24	<b>1.1.1.3 The simulation parameters</b>	<b>6</b>
25	<b>1.1.1.4 The simulation procedure</b>	<b>7</b>
26	<b>1.1.2 Experiment 2</b>	<b>9</b>
27	<b>1.1.2.1 The experimental task</b>	<b>9</b>
28	<b>1.1.2.2 The experimental procedure</b>	<b>9</b>
29	<b>1.1.2.3 The simulation parameters</b>	<b>10</b>
30	<b>1.1.2.4 The simulation procedure</b>	<b>10</b>
31	<b>1.2 The Baldwin Effect</b>	<b>11</b>
32	<b>1.2.1 The experimental task</b>	<b>11</b>
33	<b>1.2.2 The experimental procedure</b>	<b>13</b>
34	<b>1.2.3 The simulation parameters</b>	<b>13</b>
35	<b>1.2.4 The simulation procedure</b>	<b>15</b>
36	<b>1.3 The coevolution of learning and memory</b>	<b>16</b>
37	<b>1.3.1 Experiment 1</b>	<b>16</b>
38	<b>1.3.1.1 The experimental task</b>	<b>16</b>
39	<b>1.3.1.2 The experimental procedure</b>	<b>17</b>
40	<b>1.3.1.3 The simulation parameters</b>	<b>17</b>
41	<b>1.3.1.4 The simulation procedure</b>	<b>19</b>
42	<b>1.3.2 Experiment 2</b>	<b>21</b>
43	<b>1.3.2.1 The experimental task</b>	<b>21</b>
44	<b>1.3.2.2 The experimental procedure</b>	<b>21</b>
45	<b>1.3.2.3 The simulation parameters</b>	<b>22</b>
46	<b>1.3.2.4 The simulation procedure</b>	<b>22</b>
47	<b>2 Theoretical pilot work for the co-evolution of learning and memory</b>	<b>22</b>
48	<b>2.1 Simulation code</b>	<b>24</b>
49	<b>3 Analyses</b>	<b>28</b>
50	<b>3.1 The evolution of social learning</b>	<b>28</b>
51	<b>3.1.1 Experiment 1</b>	<b>28</b>
52	<b>3.1.1.1 Allele frequency</b>	<b>28</b>
53	<b>3.1.2 Experiment 2</b>	<b>29</b>
54	<b>3.1.2.1 Allele frequency</b>	<b>29</b>
55	<b>3.2 The Baldwin Effect</b>	<b>30</b>
56	<b>3.2.1 Allele frequency</b>	<b>30</b>
57	<b>3.3 The coevolution of learning and memory</b>	<b>31</b>
58	<b>3.3.1 Learning</b>	<b>31</b>
59	<b>3.3.2 Memory</b>	<b>31</b>
60		

62 **1 Full methods**

63 Here, we describe the methods of the studies in the same order in which they were  
64 presented in the main text. For each study, we describe (1) the experimental task,  
65 (2) the experimental procedure, (3) the simulation parameters (e.g., simulated  
66 genes, fitness), and (4) the simulation procedure.

67 **1.1 The evolution of social learning**

68 We carried out 218 simulations in which two genotypes evolved – an asocial  
69 genotype and a social genotype – in a changing environment. The different  
70 simulations used tasks of different difficulty and different forms of social learning.

71 **1.1.1 Experiment 1**

72 The first experiment investigated the evolution of a limited form of social learning in  
73 the context of a task with three levels of difficulty.

74 **1.1.1.1 The experimental task**

75 Participants took part in 125 trials of a numerical discrimination task. On each trial,  
76 participants were required to decide if an array of blue and yellow dots contained  
77 more blue dots or more yellow dots. The information participants received on  
78 whether the participants' genes marked them as an asocial learner or a social  
79 learner.

80 On asocial trials, participants saw the array of blue and yellow dots for 1s  
81 (Figure S2a). After which they made their decision by clicking one of two buttons  
82 labeled “more blue” and “more yellow”. In total there were 80 dots each of which

83 was positioned randomly on a rectangular grey canvas with no dots overlapping.  
84 The size of each dot was randomized such that the radii of the dots ranged from 10  
85 pixels to 20 pixels. The difficulty of each trial was determined by the number of dots  
86 of the majority color relative to the number of dots of the minority color. Using pilot  
87 studies we selected three difficulty levels: 52 vs 28 being easy, 45 vs 35 being  
88 moderate and 42 of the 38 being hard. The chosen levels of difficulty were arrived at  
89 through an experiment in which 120 participants completed 125 trials of the asocial task  
90 across different levels of difficulty. The resulting psychometric function, fit using a  
91 logistic function, was used to determine task difficulties that would produce performance  
92 close to ceiling (i.e. easy), ~85% accuracy (moderate) and ~65% accuracy (hard, Figure  
93 S1).

94 On social trials participants did not get to see the dots for themselves, rather  
95 they were informed of the decision of another participant randomly selected from  
96 the previous generation of the same simulation (Figure S2b). As with asocial trials,  
97 after receiving information participants made their decision by clicking one of two

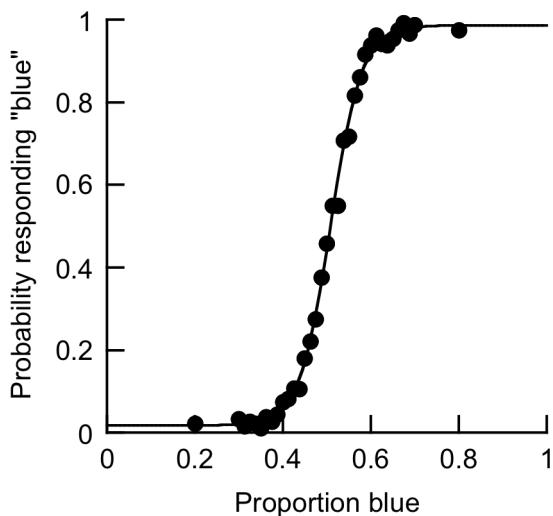


Figure S1. Results of the psychophysics experiment used to determine task difficulty.

98 buttons labeled “more blue” and “more yellow”.

99 **1.1.1.2 The experimental procedure**

100 Participants were recruited through Amazon Mechanical Turk (AMT), an online  
101 labor market where people perform short tasks for pay. Recruitment was limited to  
102 U.S. participants for whom at least 95% of their previous tasks on AMT had been  
103 approved. Recruitment and testing were approved by the Committee for Protection  
104 of Human Subjects at University of California, Berkeley and carried out in  
105 accordance with their regulations.

106 Of the 125 trials, the first 5 were practice trials. In practice trials there were 64  
107 dots in the majority so they were very easy. After this the participants took part in  
108 the remaining 120 trials in a random order. Of these 36 were easy, 36 were  
109 moderate and 36 were hard. The remaining 12 trials were “catch” trials – they were  
110 the same difficulty as the practice trials and were used to identify participants who  
111 were not paying attention to the task (see below).

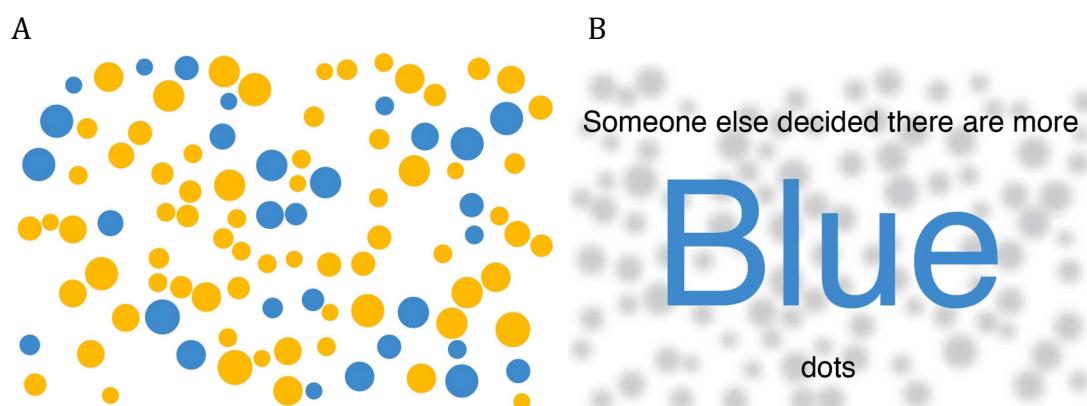


Figure S2. Stimuli from the numerical discrimination task. (a) An example of the dot arrays shown on asocial trials. In this case there are more yellow dots than blue dots and this would be considered a very easy trial. (b) An example of the social information shown on social trials.

112        For each trial, whether there were more blue or yellow dots was randomized  
113        such that each trial was independent of the others.

114        The experiment took around 10 minutes in total and participants were paid \$1  
115        for taking part. In addition, participants received a bonus of up to \$1 contingent on  
116        their success in categorizing amoeba. Performance on practice trials did not affect  
117        their bonus and participants were informed of this. Each participant's bonus was  
118        given by

$$b = \max\left(\min\left(\frac{N_s}{54} - 1, 1\right), 0\right), \quad (1)$$

119        where  $N_s$  is the number of trials (excluding practice and catch trials) on which  
120        participants made the correct decision.

121

#### 122        *1.1.1.3 The simulation parameters*

123        Each trial that participants took part in corresponded to a different agent in a  
124        different repeat simulation. Thus we were able to carry out 108 repeat simulations  
125        (excluding the 5 practice repeats and the 12 catch repeats). Within each simulation  
126        the difficulty was constant, so there were 36 hard repeats, 36 moderate repeats and  
127        36 easy repeats.

128        Each agent had a single gene with two possible values that dictated whether  
129        they were a social learner or an asocial learner. Thus, across trials, as the  
130        participants moved between different agents in different simulations whether or not  
131        they were a social learner or an asocial learner varied.

132 At the end of each trial, the fitness of the agent the participant had just played  
133 as was given by

$$f = \begin{cases} 1, & \text{asocial and right} \\ 0, & \text{asocial and wrong} \\ 1.69, & \text{social and right} \\ 0.09, & \text{social and wrong} \end{cases} \quad (2)$$

134 Each new agent inherited its gene from a single agent in the previous  
135 generation of the same simulation. The parent agent was randomly selected  
136 weighted by fitness. Inheritance was subject to mutation and with a 10% chance  
137 agents would switch from one genotype to the other. Mutation was not possible in  
138 the practice or catch simulations, thus ensuring that all these trials involved asocial  
139 learning.

140 **1.1.1.4 The simulation procedure**

141 As described above, participants took part in 125 repeat simulations. In each  
142 simulation, agents were arranged into a “discrete generational” network consisting  
143 of 40 non-overlapping generations each of 40 agents, totaling 1600 participants  
144 (Figure S3). After every 10<sup>th</sup> generation the environment changed such that the  
145 color of the majority of dots changed, but the difficulty of the task (i.e. the number of  
146 dots of the majority color) remained the same. Agents inherited their gene from an  
147 agent in the previous generation and agents in the first generation were all asocial  
148 learners (there was no previous generation from which they could learn socially).

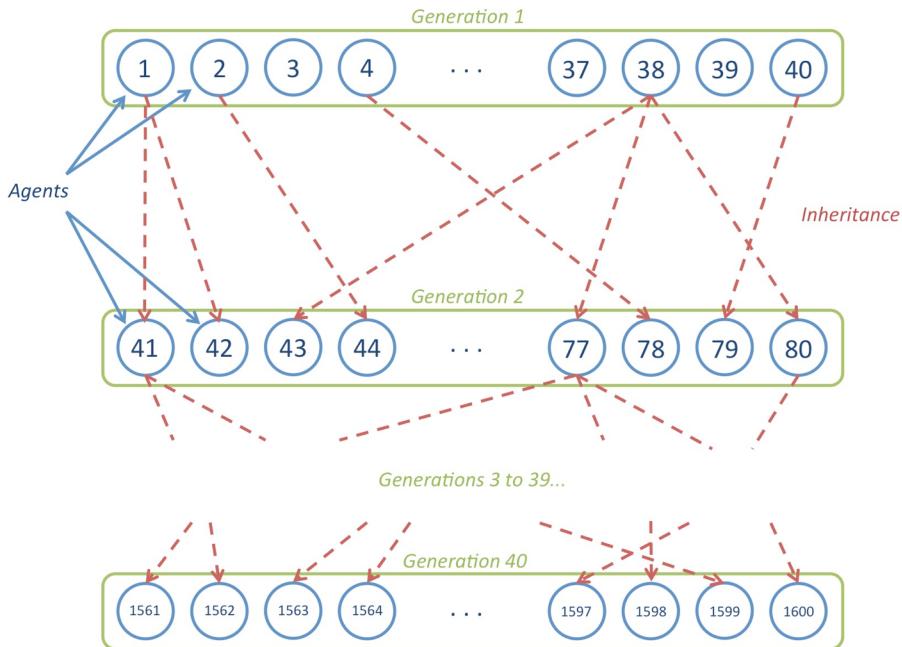


Figure S3. A discrete generational network. Agents are arranged into generations of a specific size — in this case 40. Inheritance occurs from one generation to the next. The simulation continues for a specified number of generations, here 40.

149

150        After completing all trials each participant's data was checked for  
 151        completeness and for whether the participant appeared to have attended to the  
 152        task. Participants failed the completeness check if (i) they had the incorrect number  
 153        of agents, (ii) they hadn't taken part in each repeat once, (iii) any of their agents did  
 154        not have a fitness. Participants failed the attention check if they got fewer than 10 of  
 155        the 12 catch trials correct. If either check was failed, or if the participant failed to  
 156        finish the experiment, the participant was removed from the network and another  
 157        participant was recruited to replace them.

158 In total, 1600 participants (83.7%) completed the experiment successfully, 24  
159 (1.2%) failed the attention check, 4 (0.2%) failed the data check, 232 (12.0%) quit  
160 without finishing and 53 (2.7%) ran out of time (the time limit was 30 minutes).

161 **1.1.2 Experiment 2**

162 The second experiment investigated the evolution of a more complex form of social  
163 learning. It was the same as the previous experiment with the following differences:

164 **1.1.2.1 The experimental task**

165 Participants took part in 89 trials of the numerical discrimination task. As before, on  
166 each trial, participants were required to decide if an array of blue and yellow dots  
167 contained more blue dots or more yellow dots. The information participants  
168 received on whether the participants' genes marked them as an asocial learner or a  
169 social learner.

170 On asocial trials the task was the same as before, with difficulty held constant  
171 to moderate, however, on social trials it was different. On some social trials, as in the  
172 first experiment, participants were informed of the decision of a single previous  
173 participant. However on other social trials, participants were informed of the  
174 decisions of all 40 agents in the previous generation of the same simulation (Figure  
175 S4). Moreover, in this experiment participants were informed that between every  
176 generation there was a 10% chance that the dots would change color causing the old  
177 information to be out of date.

178 **1.1.2.2 The experimental procedure**

179 As before, the first 5 trials were very easy practice trials and 12 of the subsequent  
180 trials were catch trials. Of the remaining 72 trials – 36 provided social learners with  
181 the decision of a single individual whilst the remaining 36 provided social learners  
182 with the decisions of all 40 agents from the previous generation.

In the previous batch:  
**30 people decided BLUE and 10 people decided YELLOW**

Figure S4. An example of the social information participants might receive on social trials. In this case the participant can see the decisions of all 40 participants in the previous generation, 30 of whom chose blue and 10 of whom chose yellow.

183 **1.1.2.3 The simulation parameters**

184 Each trial that participants took part in corresponded to a different agent in a  
185 different repeat simulation. Thus we were able to carry out 72 repeat simulations  
186 (excluding the 5 practice repeats and the 12 catch repeats). Within each simulation  
187 the nature of social learning was constant, so there were 36 repeats where social  
188 learning involved seeing the decision of one other individual and 36 repeats where  
189 social learners were informed of the entire previous generation.

190 **1.1.2.4 The simulation procedure**

191 In each of the repeat simulations the color of the dots changed every 10 generations,  
192 however, each simulation was staggered such that some in some simulations the  
193 dots changed color after generations 1, 11, 21 and so on, while in others the dots  
194 changed color after generations 2, 12, 22 and so on. This meant that whether or not

195 you thought the dots had recently changed color in one simulation that told you  
196 nothing about other simulations.

197 In total 1600 participants (83.0%) successfully completed the experiment, 0  
198 failed the data check, 0 failed the attention check, 259 (13.4%) quit without  
199 finishing and 69 (3.6%) ran out of time (the time limit was 30 minutes).

200 **1.2 The Baldwin Effect**

201 We carried out 18 behavioral simulations in which error-correcting genes could  
202 evolve in the context of a categorization task. The different simulations used Type I,  
203 Type II and Type V categorization rules.

204 **1.2.1 The experimental task**

205 Participants took part in 18 trials of a categorization task. On each trial, participants  
206 were required to categorize 8 different amoeba into two categories; “good” and  
207 “bad”. The amoeba varied in 3 dimensions; body color (blue or orange), nucleus  
208 color (green or purple) and spottiness (spotty or not spotty). On every trial, 4 of the  
209 amoeba were “good” and the other four were “bad” however which of the amoeba  
210 were good or bad varied across trials. On every trial the correct categorization was  
211 determined by either a Type I, Type II or Type V rule (Figure S5a-c). Type I rules  
212 depend on a single dimension, for example “blue amoeba are good”. Type II rules  
213 depend on a conjunction of two dimensions, for example “blue and spotty amoeba  
214 and orange and not spotty amoeba are good”. Type V rules are like a Type I rule, but  
215 with a pair of exceptions, for example “amoeba with green nuclei are good, unless

216 the amoeba has a blue body and spots in which case the amoeba with a purple  
217 nucleus is good and the amoeba with the green nucleus is bad".

218 At the start of each trial, participants were shown the correct categorization  
219 of all 8 amoeba for 15 seconds (Figure S5a-c). They were then sequentially shown  
220 all 8 amoeba, in a random order, and asked to categorize them by pressing "up" for  
221 good or "down" for bad (Figure S5d).

222

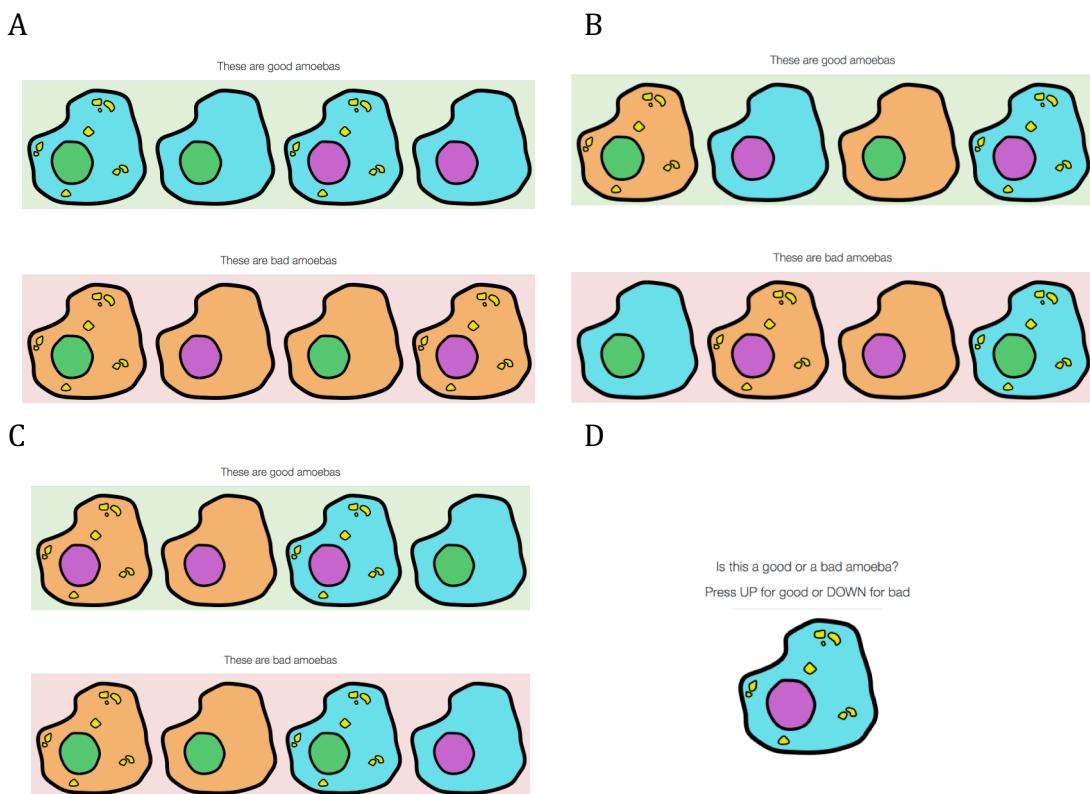


Figure S5. (a) A Type I rule, in this case the rule is "blue amoeba are good". (b) A Type II rule, in this case the rule is "orange amoeba with green nuclei and blue amoeba with purple nuclei are good". (c) A Type V rule, in this case the rule is "amoeba with purple nuclei are good, unless they are blue and don't have spots, in that case the amoeba with a green nucleus is good". (d) After being shown the correct categorization of all 8 amoebas participants were then asked to categorize all 8 amoeba themselves. In the image shown the participant is being prompted to categorize the blue spotty amoeba with a purple nucleus.

223    **1.2.2 The experimental procedure**

224    As before, participants signed up through Amazon's Mechanical Turk, gave consent,  
225    were provided with instructions, completed the trials and were then debriefed.

226    Of the 18 trials, the first three were practice trials, involving a Type I, II and V  
227    rule respectively. Of the remaining 15 trials, 3 used a Type I rule, 6 used a Type II  
228    rule and 6 used a Type V rule and participants completed them in a random order.  
229    Which dimensions were involved in the rule, and which values corresponded to  
230    "goodness" was randomly selected at the start of the simulation and varied across  
231    trials.

232    The experiment took around 10 minutes in total and participants were paid \$1  
233    for taking part. In addition, participants received a bonus of up to \$1 contingent on  
234    their success in categorizing amoeba. Performance on practice trials did not affect  
235    their bonus and participants were informed of this. Each participant's bonus was  
236    given by

$$b = \max \left( \min \left( \frac{N_s}{60} - 1, 1 \right), 0 \right), \quad (3)$$

237    where  $N_s$  is the number of amoeba participants successfully categorized.

238    **1.2.3 The simulation parameters**

239    Each trial that participants took part in corresponded to a different agent in a  
240    different repeat simulation. Thus, we were able to carry out 15 repeat simulations  
241    (excluding the 3 practice repeats). Within each simulation the type of the  
242    categorization rule was constant, so 3 of the repeats involved a Type I rule, 6  
243    involved a Type II rule and the remaining 6 involved a Type V rule.

244        Each agent had a genome consisting of 8 different genes. Each gene  
245        corresponded to one of the 8 amoeba and had two possible alleles: a neutral allele  
246        and a beneficial allele. The neutral allele had no effect on the agent, however, the  
247        beneficial allele automatically corrected categorization errors made by the  
248        participant with regards to the amoeba that the gene corresponded to. This error  
249        correction was hidden from the participant and did not affect their bonus, however  
250        it did affect the fitness of the agent.

251        At the end of each trial the fitness of the agent the participant had just played  
252        as was given by

$$f = \max \left( \left( \frac{N}{4} - 1 \right)^3, 0.0001 \right), \quad (4)$$

253        where  $N$  is the number of the 8 amoeba that were successfully categorized by the  
254        participant, or were corrected by the beneficial alleles. The fitness function is cubed  
255        to increase fitness differences, allowing selection to proceed on the relatively small  
256        population sizes we used.

257        Each new agent inherited its genes via simulated sexual reproduction from two  
258        agents in the previous generation. The parent agents were selected at random,  
259        weighted by fitness. The offspring's genome was a random combination of their  
260        parents' genes with each gene being equally likely to be inherited from one parent  
261        or the other. Inheritance was subject to mutation – for each gene there was a 5%  
262        chance its contents would be randomized producing a neutral allele with a 75%  
263        chance or a beneficial allele with a 25% chance.

264

265     **1.2.4 The simulation procedure**

266     As described above, participants took part in 18 repeat simulations. In each  
267     simulation, agents were arranged into a “discrete generational” network consisting  
268     of 40 non-overlapping generations each of 60 agents, totaling 2400 participants.  
269     Agents inherited their genes from agents in the previous generation of the same  
270     network and agents in the first generation were randomly initialized with genes  
271     reflecting the expected output of mutation: each allele of each agent had a 75%  
272     chance of having the neutral allele and a 25% chance of having the beneficial allele.

273           After completing all trials, each participant’s data was checked for  
274     completeness and for whether the participant appeared to have attended to the  
275     task. Participants failed the completeness check if (i) they had not taken part in each  
276     simulation once, (ii) any of their agents did not have a fitness, or (iii) they had not  
277     categorized the correct number of amoeba. Participants failed the attention check if  
278     they averaged less than 7 out of 8 amoeba correctly categorized on trials with a Type  
279     I rule. If either check was failed, or if the participant failed to finish the experiment,  
280     the participant was removed from the network and another participant was  
281     recruited to replace them.

282           In total, 2400 participants (72.7%) completed the experiment successfully, 460  
283     (13.9%) failed the attention check, 60 (1.8%) failed the data check, 293 (8.9%) quit  
284     early and 87 (2.6%) ran out of time (the time limit was 30 minutes).

285

286     1.3 The coevolution of learning and memory

287     1.3.1 Experiment 1

288     We carried out two simulations, one in which both learning and memory were able  
289     to evolve, and another in which only learning was able to evolve.

290     1.3.1.1 *The experimental task*

291     Participants performed 20 trials of a sequential decision-making task in the form of  
292     a “treasure hunt” game. Each trial was labeled with a country and participants were  
293     tasked with finding which of 10 locations in that country contained the treasure.  
294     Each location was represented with a labeled image and the 10 images were  
295     arranged in a ring around a central panel containing the name and flag of the  
296     country the trial was set in (Figure S6a). Before committing to a decision,  
297     participants were required to check between 1 and 10 of the locations. Checking a  
298     location informed participants whether the treasure was hidden at that location.  
299     Participants checked locations by clicking on the corresponding images, and, if the  
300     treasure was at that location, the image was replaced with an icon of a treasure  
301     chest; otherwise it was replaced with a red X (Figure S6b). After checking the  
302     required number of locations, all the location images returned to their initial state  
303     and participants were asked to guess which location had the treasure (Figure S6c).  
304     If participants found the treasure during the checking period, this decision ought to  
305     be trivial. However, if they had not found the treasure, they would need to guess by  
306     selecting among the unchecked locations.

307     Across trials, participants might revisit a country. In such cases, participants

308 were unable to check any more locations — they were required to immediately  
309 make a decision based upon their memory of their previous visit (Figure S6d).

310 **1.3.1.2 The experimental procedure**

311 Participants signed up through Amazon's Mechanical Turk, gave consent, were  
312 provided with instructions, completed the trials and were then debriefed.

313 Upon signing up, participants were asked to give their consent. If consent was  
314 given, the participants proceeded through a series of instruction pages that  
315 described the task. After the instructions, participants completed all 20 trials and  
316 were then debriefed.

317 The experiment took around 5 minutes and participants were paid \$0.50 for  
318 taking part. In addition, participants received a bonus of up to \$0.50 contingent on  
319 their success in finding the treasure. The bonus earned by each participant was  
320 given by

$$b = \max \left( \min \left( \frac{10N_s - N_c}{200}, 0.5 \right), 0 \right), \quad (5)$$

321 where  $N_s$  is the number of trials on which participants chose the correct location  
322 and  $N_c$  is the number of times participants checked a location.

323 **1.3.1.3 The simulation parameters**

324 Each participant had two simulated genes: a memory gene and a learning gene. The  
325 learning gene was an integer between 1 and 10 and determined how many locations  
326 participants could check per trial in an unfamiliar country. Thus, the number of  
327 locations participants could check was constant across trials for any given

328 participant, but varied across participants. Because a lower value of the learning  
 329 gene made the task harder, the bonuses were adjusted accordingly (Equation (5)).

330 The memory gene was a positive integer that affected the probability that a  
 331 participant would recognize a previously visited country. For each participant, at  
 332 the start of the experiment, 4 countries were chosen from a list of 75. On each trial,  
 333 one of the four was randomly selected to be the visited country. However, if the



Figure S6. Screen shots from the treasure hunt task. (a) Upon arrival at a new country participants are instructed to check a number of possible locations. (b) Checking a location reveals either the treasure or a red X. (c) After finishing checking all locations return to their original images and the participant must make a final decision. (d) If participants revisit a familiar location they cannot check locations again and were asked to make a single decision immediately.

334 participant had visited it before but the elapsed time since the previous visit  
335 exceeded the value of their memory gene, that country was replaced with another  
336 country from the list of 75 that had not yet been visited, simulating forgetting. Thus,  
337 a memory value of 0 prevents participants from recognizing previously visited  
338 countries, masking their memory. In contrast, a participant with a memory value of  
339 10 would likely recognize their repeated revisiting of a small number of countries,  
340 and so would be able to use their memory.

341 The agent's fitness was given by

$$f = (10 + 10N_s - N_c - 2M)^2, \quad (6)$$

342 where  $M$  is the value of the agent's memory gene. The fitness function is squared to  
343 increase fitness differences and allow selection to proceed more rapidly.

344 Each agent inherited genes from a randomly selected parent agent in the  
345 previous generation, weighted by fitness. Inheritance was subject to mutation —  
346 there was a 50% chance the value of each gene would be unchanged, a 25% chance  
347 that the value would increase by 1, and a 25% chance that it would decrease by 1,  
348 unless the mutation would produce a value outside the permitted range (e.g. a  
349 negative value for memory), in which case it was prevented. In the control  
350 simulation without memory, the memory genes of all participants were forced to  
351 have a value of 0.

352 **1.3.1.4 The simulation procedure**

353 In both simulations, agents were arranged into a “discrete generational” network —

354 each had 40 non-overlapping generations of 40 agents, totaling 1600 agents per  
355 simulation. Agents inherited their genes from a single agent in the previous  
356 generation. Agents in the first generation inherited their genes from a simulated  
357 agent with both learning and memory of 1.

358 After completing the experiment, each participant's data was checked for  
359 completeness and for whether the participant attended to the task. The  
360 completeness check was failed if the participant either (i) had taken part as the  
361 incorrect number of agents, (ii) had the incorrect number of genes, (iii) was  
362 connected to an incorrect number of other agents, (iv) had made an incorrect  
363 number of decisions or checks, or (v) did not have a fitness. The attention check was  
364 failed if, more than twice, the participant found the treasure but did not then choose  
365 that location as their final decision. If either check was failed, or if the participant  
366 failed to finish the experiment, the participant was removed from the network and  
367 another participant was recruited to replace them.

368 In the simulation with both memory and learning, 1600 (81.1%) participants  
369 completed the experiment successfully, 181 (9.2%) failed the attention check, 18  
370 (0.9%) failed the data check, 125 (6.3%) quit without finishing and 48 (2.4%) ran  
371 out of time (the time limit was 15 minutes). With only learning allowed to evolve,  
372 1600 (79.6%) participants completed successfully, 233 (11.6%) failed the attention  
373 check, 16 (0.8%) failed the data check, 130 (6.5%) quit without finishing and 32  
374 (1.6%) timed out.

375    **1.3.2 Experiment 2**

376    The second experiment was the same as the previous experiment with the following  
377    differences:

378    **1.3.2.1 The experimental task**

379    Participants took part in 40 rounds of the “treasure hunt” task. In this simulation,  
380    participants were allowed to check locations for treasure even on revisits to a  
381    familiar country. Moreover, the value of their learning gene no longer controlled the  
382    number of checked locations. Rather, it set the upper limit on the number of  
383    locations that could be checked — doing so was optional. Finally, the treasure  
384    changed location: at the end of each trial, the treasure changed locations with  
385    probability 0.40. The time limit was 20 minutes.

386

387    **1.3.2.2 The experimental procedure**

388    The procedure was the same as before. Participants were informed that the treasure  
389    could move. The experiment took around 7 minutes in total and participants were  
390    paid \$0.60 for taking part. In addition, participants received a bonus of up to \$0.60  
391    contingent on their success in choosing the locations where the treasure was  
392    hidden. The bonus earned by each participant was given by

$$b = \max \left( \min \left( \frac{3(10N_s - N_c - 40)}{1000}, 0.6 \right), 0.0 \right). \quad (7)$$

393     1.3.2.3 *The simulation parameters*

394     Fitness was given by

$$f = (10 + 10N_s - N_c - 4M - 4L)^2, \quad (8)$$

395     where  $L$  is the value of the agent's learning gene. The cost of memory was doubled  
396     from 2 to 4 because participants took part in twice as many trials as before. In the  
397     previous experiment checking was costly, but also mandatory, so it was equivalent  
398     to a fixed cost proportional to the value of the learning gene. Here, because checking  
399     was optional, we split the cost into both a fixed cost proportional to the value of the  
400     learning gene, plus a cost proportional to the number of time an agent checked  
401     locations. Thus, the fitness function assumes that learning has both a fixed cost,  
402     whether or not you engage in checking, as well as a direct cost for each checked  
403     location.

404

405     1.3.2.4 *The simulation procedure*

406     As before, the simulation used a discrete generational structure. We observed that  
407     1600 (82.3%) participants completed the experiment successfully, 79 (4.1%) failed  
408     the attention check, 24 (1.2%) failed the data check, 122 (6.3%) quit without  
409     finishing and 119 (6.1%) timed out.

410     **2 Theoretical pilot work for the co-evolution of learning and  
411       memory**

412     In the main paper, we presented three simulations investigating the coevolution of  
413     learning and memory. Specifically, we first considered the case of a static

414 environment, both with and without memory. We found evidence that the two  
415 would coevolve with learning reaching a higher value when memory was permitted.  
416 After this, we then sought to test the effects of environmental change on this co-  
417 evolution. The theory that these experiments were based upon suggests that  
418 environmental change would prevent the co-evolution.

419 To identify an appropriate rate of environmental change to test this  
420 hypothesis, we ran an agent-based simulation with a configurable rate of  
421 environmental change. Across repeats of the simulation, we considered three  
422 behaviors. The first matched that of the theory our experiments were based upon:  
423 agents check as many locations as they can, but upon returning to a familiar country  
424 they make the same decision as before. In addition, we considered a slightly more  
425 strategic behavior, but that still struggled to use memory effectively: agents check  
426 locations until they find the treasure or until they cannot check any more locations,  
427 upon returning to a familiar country they make the same decision as before. Finally,  
428 we considered a more complex behavior: agents check locations until they find the  
429 treasure or until they cannot check any more locations, upon returning to a familiar  
430 country they first check the location that they chose on their last visit. If the treasure  
431 is not there they continue checking locations until they find the treasure or cannot  
432 check any more locations.

433 We varied the rate of environmental change until we found a value for which  
434 these three behaviors produced reliably different evolutionary outcomes. The value  
435 we selected was 0.4. In this case, if agents perform the behavior assumed in the  
436 theory our work is based upon then neither learning nor memory evolves. If agents

437 perform the intermediate behavior then learning, but not memory evolves. Finally,  
438 the more complex strategy allowed both learning and memory to coevolve. Thus,  
439 this parameter value represents a point where the assumed behavior prevents the  
440 coevolution, but more complex behaviors may still permit it and we sought to test  
441 the effects of real human behavior.

442

443 **2.1 Simulation code**

444 The simulations were carried out in R. The code was as follows:

```
445     # number of repeats
446     n_reps <- 20
447     # number of generations per repeat
448     n_gens <- 40
449     # population size
450     N <- 40
451
452     # are the agents smart
453     # 0=basic
454     # 1=moderate
455     # 2=advanced
456     smart_agents <- 2
457
458     # number of trials in a lifetime
459     n_trials <- 40
460     # number of bandits
461     n_bandits <- 4
462     # number of arms per bandit
463     n_arms <- 10
464     # the good arms
465     good_arms <- sample(c(1:n_arms), n_bandits, replace=TRUE)
466     # number of possible decisions at each trial
467     n_pulls <- 10
468     # probability of good_arm changing
469     p_change <- 0.4
470     # probability participants correctly remember their decision
471     p_remember <- 0.8
472
473     # fitness parameters
474     f_min <- 10
475     f_pow <- 2
476
477     # payoff from getting right arm
478     payoff <- 10
479     # the cost of memory
480     c_m <- (n_trials*payoff/n_arms)*0.1
481     # the cost of curiosity
```

```

482   c_c <- (n_trials*payoff/n_arms)*0.1
483   # the cost of each check
484   c_check <- 1*payoff/n_arms
485
486   # the probability of mutation
487   q <- 0.5
488
489   # turn cognitive abilities on or off
490   allow_c <- TRUE
491   allow_m <- TRUE
492
493   # vectors to store the data
494   m_results <- array(1, dim=c(n_reps, n_gens))
495   c_results <- array(1, dim=c(n_reps, n_gens))
496
497   # for every repeat
498   for (rep in 1:n_reps) {
499     # reset genes
500     if (allow_m == TRUE) {
501       M <- rep(1, N)
502     } else {
503       M <- rep(0, N)
504     }
505     if (allow_c == TRUE) {
506       C <- rep(1, N)
507     } else {
508       C <- rep(0, N)
509     }
510
511     # for every generation
512     for (gen in 1:n_gens) {
513       #reset fitness
514       f <- rep(0, N)
515
516       # for every individual
517       for (i in 1:N) {
518
519         # initialize data vectors
520         success <- rep(FALSE, n_trials)
521         num_checks <- rep(0, n_trials)
522         decisions <- rep(0, n_trials)
523
524         # pick the bandits they visit:
525         bandits <- sample(c(1:n_bandits), n_trials, replace=TRUE)
526
527         # for every trial
528         for (t in 1:n_trials) {
529
530           # give the bandits a chance to change their good arm
531           for (b in 1:n_bandits) {
532             if (runif(1, 0, 1) < p_change) {
533               good_arms[b] <- sample(c(1:n_arms), 1)
534             }
535           }
536
537           # get their memory of bandits, decisions and successes
538           remember_bandit <- FALSE
539           if (M[i] > 0 & t > 1) {

```

```

540     memory_start <- max(1, t-M[i])
541     memory_end <- t-1
542
543     remembered_bandits <- bandits[memory_start:memory_end]
544     remembered_decisions <-
545     decisions[memory_start:memory_end]
546         remembered_successes <-
547     success[memory_start:memory_end]
548
549     remember_bandit <- bandits[t] %in% remembered_bandits
550     if (remember_bandit == TRUE) {
551         if (runif(1, 0, 1) < p_remember) {
552             remembered_decision <-
553             tail(remembered_decisions[remembered_bandits == bandits[t]], 1)
554         } else {
555             remembered_decision <- sample(c(1:n_arms), 1)
556         }
557     }
558
559     if (smart_agents == 2) {
560         # if the agents are smart
561         # if they remember the bandit, they check their past
562         decision
563             if (remember_bandit == TRUE & remembered_decision ==
564             good_arms[bandits[t]]) {
565                 decisions[t] <- remembered_decision
566                 num_checks[t] <- 1
567             } else {
568                 # otherwise check arms
569                 checked_arms <- sample(c(1:n_arms), C[i],
570 replace=FALSE)
571
572                 if (good_arms[bandits[t]] %in% checked_arms) {
573                     decisions[t] <- good_arms[bandits[t]]
574                     num_checks[t] <- match(good_arms[bandits[t]],
575 checked_arms)
576                     } else {
577                         decisions[t] <- sample(c(1:n_arms)[!c(1:n_arms)
578 %in% checked_arms], 1)
579                         num_checks[t] <- C[i]
580                     }
581                 }
582             }
583         } else if (smart_agents == 1) {
584             # if the agents are moderate
585             if (remember_bandit == TRUE) {
586                 # if they remembered the bandit
587                 # do what they did last time
588                 decisions[t] <- remembered_decision
589                 num_checks[t] <- 0
590             } else {
591                 # otherwise check arms
592                 checked_arms <- sample(c(1:n_arms), C[i],
593 replace=FALSE)
594
595                 if (good_arms[bandits[t]] %in% checked_arms) {
596                     decisions[t] <- good_arms[bandits[t]]

```

```

597             num_checks[t] <- match(good_arms[bandits[t]],
598 checked_arms)
599             } else {
600                 decisions[t] <- sample(c(1:n_arms)[!c(1:n_arms)
601 %in% checked_arms], 1)
602                     num_checks[t] <- C[i]
603                     }
604             }
605         } else {
606             if (remember_bandit == TRUE) {
607                 # if they remembered the bandit
608                 # do what they did last time
609                 decisions[t] <- remembered_decision
610                 num_checks[t] <- 0
611             } else {
612                 # otherwise check arms
613                 checked_arms <- sample(c(1:n_arms), C[i],
614 replace=FALSE)
615
616                 if (good_arms[bandits[t]] %in% checked_arms) {
617                     decisions[t] <- good_arms[bandits[t]]
618                     num_checks[t] <- C[i]
619                 } else {
620                     decisions[t] <- sample(c(1:n_arms)[!c(1:n_arms)
621 %in% checked_arms], 1)
622                         num_checks[t] <- C[i]
623                         }
624                 }
625             }
626             success[t] <- (decisions[t] == good_arms[bandits[t]])
627         }
628
629         #calculate fitness
630         f[i] <- max(f_min + sum(success)*payoff -
631             sum(num_checks)*c_check - C[i]*c_c - M[i]*c_m, 0.001)
632             #f[i] <- max(f_min + sum(success)*payoff - sum(num_checks)
633             - M[i]*c_m, 0.001)
634         }
635
636         # assign fitness as a probability
637         fb <- (f/100)**f_pow
638         f2 <- fb/sum(fb)
639
640         # save state
641         m_results[rep, gen] <- mean(M)
642         c_results[rep, gen] <- mean(C)
643
644         #do reproduction and mutation
645         M2 <- M
646         C2 <- C
647         fdum <- cumsum(f2)
648         parents <- runif(N,0,1)
649         for (i in 1:N) {
650             parents[i] <- match(TRUE, fdum>parents[i])
651         }
652         if (allow_m == TRUE) {
653             M2 <- pmax(M[parents] + sample(c(0, 1, -1), N, prob=c(1-q,
654 q/2, q/2), replace=TRUE), 0)

```

```

655      }
656      if (allow_c == TRUE) {
657        C2 <- pmin(pmax(C[parents] + sample(c(0, 1, -1), N,
658        prob=c(1-q, q/2, q/2), replace=TRUE), 1), n_pulls)
659      }
660      M <- M2
661      C <- C2
662    }
663  }

```

664 **3 Analyses**

665 We analyzed the data with Bayesian models using MCMC methods to estimate  
 666 parameter values. All analyses were carried using the `rjags` package in R. Unless  
 667 otherwise stated, all parameter estimates are based on >3000 independent samples  
 668 generated from 3 chains. The values quoted in the main paper are the central  
 669 credible intervals of these samples presented as the median sample and the 2.5%  
 670 and 97.5% quantiles. In all cases, we used vague priors.

671 **3.1 The evolution of social learning**

672 **3.1.1 Experiment 1**

673 **3.1.1.1 Allele frequency**

674 We modeled the probability an agent was a social learner as a Bernoulli variable  
 675 with a logit link function. The linear predictor contained a fixed effect for the  
 676 number of generations since the environment changed (ranging from 0 to 9) for  
 677 each of the three levels of difficulty and an additional random effect for the repeat,  
 678 such that:

$$social\ learner \sim Bern(p) \quad (9)$$

679 where:

$$\text{logit}(p) = \beta_{g,d} + \varepsilon_r \quad (10)$$

680 where  $\beta$  and  $\varepsilon$  are estimated effect parameters,  $g$  is the number of generations since  
681 environmental change,  $d$  is the difficulty level and  $r$  is repeat number.

682 The priors were as follows:

$$\beta \sim N(0.0, 0.01) \quad (11)$$

683

$$\varepsilon \sim N(0.0, \tau) \quad (12)$$

684

$$\tau \sim \text{gamma}(0.001, 0.001) \quad (13)$$

685 Data from the practice trials was discarded, as was data from the first 10  
686 generations, to allow the population to reach equilibrium.

687 The average frequency of social learning for each difficulty level was calculated  
688 by averaging the  $\beta$  parameters for that level of difficulty. The magnitude of the drop  
689 in social learning following environmental change was calculated by taking the  
690 difference between  $\beta_{1,d}$  and  $\beta_{3,d}$ .

691

### 692 3.1.2 Experiment 2

#### 693 3.1.2.1 Allele frequency

694 The analysis was as described in section 3.1.1.1, except that the social information  
695 condition took the place of difficulty.

696    

## 3.2 The Baldwin Effect

697    

### 3.2.1 Allele frequency

698    We modeled the probability that an allele present in an agent was beneficial as a  
699    Bernoulli variable with a logit link function. The linear predictor contained a fixed  
700    effect for the type of learning rule the agent was learning, an effect of whether the  
701    amoeba that the gene corresponded to was an exception and a random effect for  
702    repeats. Such that:

$$\text{allele} \sim \text{Bern}(p) \quad (14)$$

703    where:

$$\text{logit}(p) = \begin{cases} \beta_t + \varepsilon_r, & \text{nonexception} \\ \beta_t + \gamma + \varepsilon_r, & \text{exception} \end{cases} \quad (15)$$

704    where  $\beta$ ,  $\gamma$  and  $\varepsilon$  are estimated parameters,  $t$  is the type of rule and  $r$  is the repeat  
705    number.

706    The priors were as follows:

$$\beta \sim N(0.0, 0.001) \quad (16)$$

707

$$\varepsilon \sim N(0.0, \tau) \quad (17)$$

708

$$\tau \sim \text{gamma}(0.001, 0.001) \quad (18)$$

709

710 Data from the practice trials was discarded as was all data from the first 39  
711 generations as we wished to study the frequency of the beneficial allele in the final  
712 generation.

713 **3.3 The coevolution of learning and memory**

714 **3.3.1 Learning**

715 We modeled the value of an agent's learning gene as a normally distributed variable.  
716 The linear predictor contained a fixed effect for the generation the agent was in.  
717 Such that:

$$\text{learning} \sim N(\mu, \sigma^2) \quad (19)$$

718 where:

$$\mu = \beta_g \quad (20)$$

719 where  $\beta$  is a parameter to be estimated and  $g$  is the generation within the  
720 simulation.

721 The priors were as follows:

$$\beta \sim N(0.0, 0.001) \quad (21)$$

722

$$\sigma^2 \sim \text{gamma}(0.001, 0.001) \quad (22)$$

723 **3.3.2 Memory**

724 Memory was analysed in the same way as learning (see section 3.3.1) such that:

$$memory \sim N(\mu, \sigma^2) \quad (23)$$

725 where:

$$\mu = \beta_g \quad (24)$$

726 where  $\beta$  is a parameter to be estimated and  $g$  is the generation within the  
727 simulation.

728 The priors were as follows:

$$\beta \sim N(0.0, 0.001) \quad (25)$$

729

$$\sigma^2 \sim gamma(0.001, 0.001) \quad (26)$$

730

731