

1 **How cognitive and environmental constraints influence**
2 **the reliability of simulated animats in groups**

3

4 **Dominik Fischer**^{1, #a, *}, **Sanaz Mostaghim**³, and **Larissa Albantakis**^{2, §}

5

6 ¹ School of Management, Technical University of Munich; d.fischer@tum.de

7 ² Department of Psychiatry, University of Wisconsin–Madison; albantakis@wisc.edu

8 ³ Faculty of Computer Science, Otto von Guericke University of Magdeburg;

9 sanaz.mostaghim@ovgu.de

10

11 * Correspondence: d.fischer@tum.de; Tel.: +49-89-289-23239

12 #a Current Address: School of Management, Technical University of Munich, Germany

13 § Funding sources: L.A. receives funding from the Templeton World Charities Foundation

14 (Grant #TWCF0196).

15

Abstract

16 Evolving in groups can either enhance or reduce an individual’s task performance. Still,
17 we know little about the factors underlying group performance, which may be reduced to
18 three major dimensions: (a) the individual’s ability to perform a task, (b) the dependency on
19 environmental conditions, and (c) the perception of, and the reaction to, other group
20 members. In our research, we investigated how these dimensions interrelate in simulated
21 evolution experiments using adaptive agents equipped with Markov brains (“animats”). We
22 evolved the animats to perform a spatial-navigation task under various evolutionary setups.
23 The last generation of each evolution simulation was tested across modified conditions to
24 evaluate and compare the animats’ reliability when faced with change. Moreover, the
25 complexity of the evolved Markov brains was assessed based on measures of information
26 integration. We found that, under the right conditions, specialized animats were as reliable
27 as animats already evolved for the modified tasks, that interaction between animats was
28 dependent on the environment and on the design of the animats, and that the task difficulty
29 influenced the correlation between the performance of the animat and its brain complexity.
30 Generally, our results suggest that the interrelation between the aforementioned dimensions
31 is complex and their contribution to the group’s task performance, reliability, and brain
32 complexity varies, which points to further dependencies. Still, our study reveals that
33 balancing the group size and individual cognitive abilities prevents over-specialization and
34 can help to evolve better reliability under unknown environmental situations.

35 **Keywords:** Collective behavior, evolutionary algorithms, cognitive science, Markov brains.

36

Author Summary

37 The ability to adapt to environmental changes is an essential attribute of organisms which
38 have had evolutionary success. We designed a simulated evolution experiment to better
39 understand the relevant features of such organisms and the conditions under which they
40 evolve: First, we created diverse groups of cognitive systems by evolving simulated
41 organisms (“animats”) acting in groups on a spatial-navigation task. Second, we post-
42 evolutionary tested the final evolved animats in new environments—not encountered before—
43 in order to test their reliability when faced with change. Our results imply that the ability to
44 generalize to environments with changing task demands can have complex dependencies on
45 the cognitive design and sensor configuration of the organism itself, as well as its social or
46 environmental conditions.

47 Introduction

48 *Intelligence is the ability to adapt to changes.* According to this prevalent perspective,
49 possessing general intelligence [1,2] not only enables one to perform a task correctly under
50 already known conditions, but also to perform well under unexpected conditions. Further, in
51 natural environments intelligent behavior is not only dependent on the (maybe limited)
52 intelligence of the individual organism, but also involves interactions with the social and
53 physical environment [3–5]. In addition to the examples from the animal world, it is also true
54 in *high-reliability organizations* (e.g., aircraft carrier or nuclear power plants) that individual
55 behavior is interrelated with the behavior of the group members. This is necessary to be able
56 to act correctly in case of an unforeseen event [6–8].

57 While it seems intuitive that there is a triangular relationship between the individual, the
58 group, and the environment, we discovered a lack of research on how individual behavior
59 and group behavior are interrelated and depend on spatial attributes of the environment [9].

60 This limits our understanding of how an individual actor evolves intelligent behavior and
61 how its physiological abilities, the social setting, and the environment constrain this
62 evolution. More generally, several studies have investigated intelligence and knowledge on
63 the group level, and some have modelled groups of individuals as single agents (e.g., [10–
64 14]). These studies have their origins in a variety of disciplines and have in common that they
65 seek to elucidate the dynamics between group members.

66 To shed more light on the above-mentioned issue, we wanted to ask which conditions
67 can promote the evolution of intelligent entities that act in organized groups and can
68 additionally adapt to environmental changes under simplified conditions in a simulated
69 evolution experiment. Inspired and motivated by Pinter-Wollman et al. [9], we wanted to
70 investigate how the behavior and performance of evolved “animats” (simulated agents with
71 cognitive abilities [15,16]) varies in different task conditions, such as changes in the
72 proportions of static objects, dynamic objects (group members), and individual cognitive
73 abilities. This simulation enabled us to manipulate and observe three dimensions which might
74 influence task performance and reliability: the group size, the animats’ physiology, and the
75 environmental design. In this study, reliability describes the ability to perform well under
76 manipulated task conditions that the animat had not been confronted with before.

77 We used a genetic algorithm to let the animats’ behavior evolve under various
78 evolutionary setups. Specifically, the animats were controlled by *Markov brains (MBs)* [16],
79 which consisted of computational units whose functions and connectivity were determined
80 by the animats’ adaptive genome. The animats’ task was to navigate through a two-
81 dimensional world composed of multiple rooms without colliding with other group members
82 (see Fig 1). There was a small penalty for each collision and a large reward for crossing gates
83 between rooms. After an evolution of 10,000 generations, we tested the final animats under
84 modified task conditions modeled as: a variation in group size, the complexity of the static

85 obstacles in the environment, and interaction rules between animats constraining fitness for
86 the task. An animat was considered reliable if its task performance remained high across
87 many of these test conditions.

88 **Fig 1. Average movement patterns of six selected conditions.** The panel on the left shows the two-
89 dimensional environment including two rooms with 36 start positions occupied (round dots). The
90 other six panels show example movement patterns. Dark fields indicate high occupancy, and light
91 fields indicate low occupancy in the corresponding position throughout the trial. Generally, well-
92 performing animat groups evolve a wall following strategy. $\langle EF \rangle$ indicates the average fitness of
93 the final generation in the specific condition.

94 A predecessor study focused on the influence of group size on the evolution of group
95 fitness and reliability [17], while the present work extends the reliability experiments,
96 includes cognitive and environmental variations in the evolutionary setup, and elaborates the
97 measurement of brain complexity by applying measures developed within the framework of
98 the integrated information theory (IIT) to the evolved MBs [18,19]. There are two additional
99 works which directly relate to our study: First, Konig et al. [20] provided the original
100 experimental setup. They designed a two-dimensional spatial-navigation task in which a
101 swarm of robots has to learn to travel between two rooms. Second, Albantakis et al. [19]
102 showed how single animats evolve in a perceptual-categorization task environment with
103 dynamic objects under various task difficulties. The primary motivation behind their work
104 was to investigate the evolution of integrated information [18], which is an indicator for brain
105 complexity, and its relation to task difficulty and memory capacity. In the following, we
106 discuss how the complexity of the MBs—evolved in the various experimental setups—is related
107 to reliability as an indicator for general intelligence.

108 Simulating a large set of evolutionary setups and post-evolutionary test conditions
109 enables us to identify important cognitive and social variables and to evaluate how physical
110 constraints influence collective movement. Specifically, the results of the simulated

111 evolution experiments suggest the following implications: First, animats who evolve in an
112 environment with a balanced group size evolve better reliability and can compete with
113 specialized animats (who have already experienced changing conditions). Second, the
114 integration of motor units into the memory network increases the performance of animats.
115 Third, the ability to sense adjacent animats is essential for the reliability of animats to perform
116 the task, even if it is challenging to make statements about the communication between
117 animats in this setting. Finally, we explored how various sensor configurations influence the
118 difficulty of dealing with the task and, therefore, the animats' ability to cope with changes.
119 Overall, we found that, under the right conditions, specialized animats can be reliable, that
120 the integration of motor units has an impact on performance and reliability, that animats
121 benefit from passive interaction, and that more sensors enable reliability with simpler and
122 less integrated brain structures (which challenges the view that higher generalized
123 intelligence is necessarily associated with more complex cognitive architectures). On the
124 whole, our approach also highlights the complexity of the dependencies between the three
125 dimensions under investigation (properties of the individual, group interaction, and
126 environmental design), even in the simplified conditions of our simulation experiments, and
127 thus cautions against hasty generalizations, e.g., across different species or environments.

128 In the following, we will first present our results on the animats' task performance,
129 reliability, behavior, and brain complexity across varying evolutionary setups. After that, we
130 will discuss the findings in the broader scope of the literature and also how our work
131 contributes to it. The last part of the work explains the methods and research design.

132 **Results**

133 We simulated the evolution of artificial organisms (“animats”) with diverse cognitive
134 architectures under various conditions for 10,000 generations (see Table 1 for an overview
135 of all evolution simulations conducted).

136
137

Table 1. Definition of simulation conditions (“evolutionary setups”). G_i indicates the group condition. The index i specifies the respective evolutionary setup.

Label	G_i	Group Size ¹	Cognitive Design ²	Interaction Condition ³	Sensor Configuration ²	Results in Figures
Varying group size	<i>1.00</i>	72	4 hidden 2 motors with motors	Active Penalty, blocking disabled	1 animat sensor, 1 wall sensor	2/3/4
	<i>0.75</i>	54				
	<i>0.50</i>	36				
	<i>0.25</i>	18				
	<i>single</i> <i>random</i>	1 random				
Varying cognitive design	<i>bigbrain</i> <i>smallbrain</i> <i>!feedback</i>	36	8 hidden	Active Penalty, blocking disabled	1 animat sensor, 1 wall sensor	5/6/7
			2 hidden			
			4 hidden 2 motors wo/ feedback			
Varying interaction conditions	<i>!penalty</i> <i>blocked/ !penalty</i> <i>blocked</i>	36	4 hidden 2 motors with motors	No Penalty, blocking disabled	1 animat sensor, 1 wall sensor	8/9/10
				No penalty, blocking enabled		
				Active Penalty, blocking disabled		
Varying sensor configuration	<i>!agent</i> <i>3sides</i> <i>w=a</i>	36	4 hidden 2 motors with motors	Active Penalty, blocking disabled	1 wall sensor	11/12/13
					3 animat sensors, 3 wall sensors	
					1 universal sensor	

138
139
140
141

¹ Absolute group size, 72 animats corresponds to 100% coverage of available starting slots.

² See Methods section for detailed architecture.

³ If penalty is active, animats receive penalty for colliding with other animats. If blocking is active, animats are not able to share the same position, otherwise they can cross over each other.

142
143
144
145
146
147
148
149

All animats were evolved to travel between two rooms in a two-dimensional environment, which they shared with other animats of their same type, except in the “single” condition (see Fig 1(a) and Table 1). Fitness selection positively depended on the average number of times that the animats stepped through the gate between the two rooms. In addition, we imposed a small penalty each time they collided with other animats (if not stated otherwise). A detailed description of the task environments and the *evolutionary algorithm* (*EA*) is provided below in the Methods section. In many evolutionary setups, high final fitness values ($EF > 3$) was able to be achieved.

150
151
152
153

Once evolved, the final generation of animats was the basis for comparing task fitness (performance in a specific environment), behavior, and reliability (average performance across all several task environments) across conditions. In this study, we focused on assessing reliability across two dimensions: (1) the number of co-existing animats and (2) the

154 placement of static obstacles compared to the original two-dimensional environment (see Fig
155 1(a), and the Methods section for details). Additionally, we varied the interaction conditions
156 between agents as a third parameter to manipulate the agent's reliability across group sizes.

157 Fig 1(b) displays six different heatmaps visualizing several evolved movement patterns.
158 It is observable that animat groups with reasonable task fitness (TF) converge towards a
159 "swarm"-like wall-following behavior, which is driven by both interactions with fellow
160 animats and interactions with the environment [4,9].

161 We organized the presentation of our results into four sections according to the
162 evolutionary setups thereof, as shown in Table 1 (varying "group size", "cognitive design",
163 "interaction conditions", and "sensor configuration", respectively). Each section contains
164 visualizations displaying the average increase in fitness across generations ("fitness
165 evolution"), behavioral features, the reliability tests, and a complexity analysis of the evolved
166 MBs. Since the figures are redundant in their construction, we will briefly introduce their
167 attributes:

168 **Fitness:** Fig 2, Fig 5, Fig 8, and Fig 11 show (a) the fitness evolution across generations
169 and (b) the distribution of evolved fitness values (EF) of the final generation. The shaded
170 areas in (a) visualize the *standard error of the mean (SEM)* across the 30 evolution
171 simulations that we performed per evolutionary setup. The wide bars in (b) visualize the mean
172 evolved fitness (EF).

173 **Reliability and behavior:** Fig 3, Fig 6, Fig 9, and Fig 12 visualize the results of testing
174 the reliability of fitness values and behavioral features of the final generation of animats
175 across (1) different group sizes ($[0.01389, 0.05, 0.10, \dots, 0.95, 1.0]$) and (2) various test
176 conditions (changing interactions between animats and environment design). Panel (a) in
177 Figures 3/6/9/12, shows the mean task fitness (TF) of testing the animats under different
178 group sizes, respectively, in the eight different test conditions listed in Table 2. Note that the

179 condition under which a group of animats evolved is indicated by their G_i label (see Table
 180 1), and $\langle TF \rangle$ is an average across the 30 evolution simulations per experimental setup.

181 **Table 2:** Overview of the eight environments in which reliability tests were performed. They differ
 182 in environmental conditions and in the complexity of the world design.

Label	Environmental Conditions	Environment (see Methods)
Original	Active penalty ¹ , no blocking ²	See Fig 16(a)
No Penalty	No penalty, no blocking	See Fig 16(a)
Blocked	Active penalty, active blocking	
Blocked and no Penalty	No penalty, active blocking	
Noisy Corners Small Gates 4 Rooms 4 Messy Rooms	Active penalty, no blocking	See Fig 16(b) See Fig 16(c) See Fig 16(d) See Fig 16(e)

183 ¹ If penalty is active, animats receive penalty when colliding into each other.

184 ² If blocking is active, an animat cannot move onto the location of another animat.

185 Next, we quantified the reliability across group sizes as the average task fitness $R =$
 186 $\langle TF \rangle_{GS}$ in the “Original” test condition (in this case, the average is calculated across group
 187 sizes not simulations as indicated by the subscript “GS”, which stands for group size). Panel
 188 (b) shows the distribution of these reliability values (R) and their dependency on evolved
 189 fitness (EF). Panel (c) shows how the animats’ behavior depends on the relative group size,
 190 evaluating the probability of an animat to stand still (“no movement”), turn, or move forward
 191 in the “original” test environment.

192 **Complexity analysis:** Fig 4, Fig 7, Fig 10, and Fig 13 show two types of metrics for MB
 193 complexity: (a) the distribution of *integrated information* (Φ^{Max}) [18,19], and (b) the
 194 corresponding *number of concepts* ($\#Concepts(\Phi^{Max})$) [18] per condition. While there may
 195 be simpler, less computationally demanding options for evaluating the causal complexity of
 196 the evolved MBs (see [15,16,21]), the chosen measures are fairly well established [15,19,22]
 197 and are theoretically motivated as part of the formal framework of integrated information
 198 theory (IIT) [18]. Briefly, a “concept” in IIT is a system subset that has a causal role within
 199 the system—an intrinsic mechanism. A concept causally constraints both the past and future
 200 states of the system, and is irreducible to its parts. The number of concepts ($\#Concepts(\Phi^{Max}$
 201 $)$) thus captures the number of internal functions performed by individual system elements

202 and combinations of elements. Φ^{Max} quantifies how much of the information specified by all
203 the concepts in a set of elements would be lost under a partition of the system, and it will be
204 high if the set of elements has many concepts (functional differentiation) that are also highly
205 integrated. Both measures are evaluated for the most integrated system subset, thus the ‘max’
206 superscript. For details please refer to the original publication [18] and to [19] for an
207 application of these measures to evolve MBs.

208 **Varying group size: Evolution under specialized conditions can produce reliable** 209 **agents**

210 In a first set of experiments, we compared animats that evolved within groups of fixed
211 sizes (1-72 animats) (using the original animat and environment design in all cases).
212 Preliminary results, including a comparison of the reliability of evolution conditions $G_{1,0}$ -
213 *single*, were presented in [17]. As shown in Fig 2(a) and reported in [17], group size during
214 evolution does impact the animats’ ability to perform the gate crossing task (“task difficulty”)
215 (Fig 1(a)), and it influences the final evolved fitness.

216 In our spatial-navigation task, single animats (group size of 1) frequently find an optimal
217 solution within 10,000 generations, since colliding is impossible and walls (static obstacles)
218 may guide the animat towards the gate. Increasing the number of animats in the environment
219 makes it more difficult to navigate due to the penalty imposed upon colliding with another
220 agent [17]. In our study, an animat was reliable if it could achieve high fitness under various
221 conditions which they did not face during evolution. Reliability across group sizes was found
222 to be high if the animats evolved in an environment where the density of animats was
223 balanced ($G_{0.5}$ and $G_{0.25}$) (see Fig 3(a,b) and [17]).

224 In our study, we included an additional comparison setup (G_{random}), for which group size
225 varied randomly during evolution, in order to explicitly evolve animats with high reliability.
226 As shown in Fig 2(b), the final fitness values for G_{random} were comparable to those evolution

227 setups with fixed, intermediate group sizes ($G_{0.5}$ and $G_{0.25}$) (though still significantly different
228 ($p < .05$), see Supporting Information S3 for all statistical tests).

229 **Fig 2. Fitness evolution and distribution of the final evolved fitness.** (a) G_{single} is the condition
230 which evolves the highest fitness on average. Larger group sizes during evolution impede the
231 animats' fitness evolution and lead to lower final evolved fitness values. (b) The evolutionary setup
232 with randomized group sizes at each generation (G_{random}) demonstrates similar properties as those
233 setups with fixed, intermediate group sizes ($G_{0.25}$ and $G_{0.5}$).

234 It is no surprise that G_{random} is the most reliable setup across varying group sizes (see Fig
235 3), since these animats already evolved under the conditions tested in the reliability
236 evaluations. Notably, however, animats that evolved under specialized conditions with
237 intermediate group sizes ($G_{0.5}$ and $G_{0.25}$) are comparable to animats specifically evolved for
238 reliability (G_{random}) during evolution (see Fig 3). It is necessary to review the reliability tests
239 in detail to observe differences between those evolutionary setups. $G_{0.50}$ and G_{random} show
240 similar reliability values R in the original environment setting, particularly for larger group
241 sizes ($> 50\%$) (see Fig 3(a)). Nevertheless, G_{random} animats perform better with smaller group
242 sizes, leading to comparable but still significantly different average R values ($p < .05$).

243 **Fig 3. Reliability tests.** (a) Overall, only G_{single} fails to generalize across group sizes, as animats
244 evolved without other group members did not develop strategies to avoid collisions (compare
245 *Original* to *No penalty* test condition, where G_{single} performs well throughout). There is a large
246 difference in the *Blocked* environment between G_{random} , $G_{0.25}$, and $G_{0.50}$, while in other environments
247 their task fitness is comparable, pointing to somewhat different navigation strategies. (b) On average,
248 G_{random} is the most reliable condition, followed by $G_{0.50}$ and $G_{0.25}$. Except for G_{single} , EF correlates
249 with R in all groups. (c) Note that $G_{0.50}$ and $G_{0.25}$ change their behavior more with increasing animat
250 density compared to G_{random} .

251 Of all test conditions (see Table 2), *Blocked* (in which animats cannot overlap) suggests
252 a further difference between $G_{0.50}$, $G_{0.25}$, and G_{random} (see Fig 3(a)): $G_{0.50}$ and $G_{0.25}$ are more
253 severely affected by this deviation from standard settings in which animats can overlap, albeit

254 under a penalty. While animats evolved in G_{random} also experienced large group sizes with a
 255 higher likelihood of a penalty during evolution, $G_{0.50}$ and $G_{0.25}$ animats consistently faced
 256 only intermediate probabilities of colliding with other animats, which may have led to less
 257 effective strategies for avoiding collisions.

258 In addition to varying group sizes, we also tested the final generation of animats in four
 259 environments with different wall arrangements (Fig 3(a), bottom row). Performance
 260 decreased to similarly low levels in all conditions, but least for evolutionary setups with
 261 larger group sizes.

262 In terms of their behavior (Fig 3(c)), animats in G_{random} were less idle and showed fewer
 263 turns and more steps forward in comparison with animats in $G_{0.50}$, particularly for large group
 264 sizes. This suggests that the behavior in G_{random} is more fluid overall. By contrast, the
 265 specialized animats have to be more reactive to stay reliable, displaying larger difference in
 266 behavior across group sizes (see Table 3 for a more detailed explanation of the difference in
 267 behavior). Please refer to [17] for a more detailed discussion of behavioral differences across
 268 evolutionary setups with fixed group sizes $G_{1.0-single}$.

269 **Table 3:** Absolute difference between the state transition probability of $G_{0.50}$ and G_{random} . The first
 270 digit describes whether anything (wall or other animat) is sensed (1) or not sensed (0), and the
 271 second digit describes whether the animat moved/turned (1) or did not move/turn (0). Most notably,
 272 G_{random} animats performed more movements even in the absence of sensor inputs than $G_{0.50}$
 273 (“01→01”).

$t \backslash SM$	00	01	10	11
00	0.0000	-0.0074	0.0000	-0.0001
01	-0.0079	-0.0606 ¹	0.0136	0.0088
10	0.0005	0.0100	0.0063	0.0063
11	-0.0001	0.0119	0.0031	0.0157

274 ¹ Minus values indicate that the transition is more frequent in G_{random} , while positive values indicate the
 275 opposite.

276 Fig 4 shows the distribution of Φ^{Max} and $\#Concepts(\Phi^{Max})$ [18,19] as a measure of the
 277 complexity of the evolved MBs across evolutionary setups with different group sizes. While
 278 the most reliable evolutionary setups (G_{random} and $G_{0.50}$) do show the highest average values

279 of Φ^{Max} and the largest number of concepts (internal mechanisms), differences between
280 conditions generally do not reach statistical significance ($p > .05$) due to the large variance in
281 the complexity values (see Supporting Information S3). It would require more data
282 (simulation experiments per evolutionary setup) to refine the mean of the intervals enough to
283 verify the observed trend. In our predecessor study [17], a correlation of high reliability and
284 task performance with high brain complexity was found using a simplified measure of brain
285 complexity based on anatomical connectivity only. In addition, the integrated information
286 measures employed here are sensitive to the causal interactions within the MBs. In the
287 present data, significant pair-wise differences could be found between G_{single} and the most
288 reliable setups (G_{random} and $G_{0.50}$). As explained above, the task environment experienced by
289 animats in G_{single} is less demanding than for setups with larger group sizes. Our findings are
290 thus in line with [19], which demonstrated higher Φ^{Max} and $\#Concepts(\Phi^{Max})$ for animats
291 evolved in more complex environments.

292 **Fig 4. Distribution of brain complexity measures.** Differences in (a) Φ^{Max} and (b) the
293 corresponding number of concepts was found between the most (G_{random} and $G_{0.50}$) and the least
294 (G_{single}) reliable setups. Due to the large variance in the data and the low sample size (30 simulations
295 per evolutionary setup), differences in the mean between the remaining conditions did not reach
296 statistical significance (see Supporting Information S3).

297 **Varying cognitive design: Brain size and memory dependencies**

298 In a second set of experiments, we used the same evolutionary setup as for $G_{0.50}$ in all
299 tested conditions, but varied the number of available computational units in the animats'
300 MBs. In the baseline design $G_{0.50}$, it is possible to integrate motor units as memory units (by
301 feedback loops to the hidden units, see Methods section). This was disabled in one condition
302 $G_{feedback}$ and therefore reduced the absolute capacity for memory from six to four binary
303 units. Moreover, we designed animats with similarly small memory capacity but with
304 feedback motors as a reference group ($G_{smallbrain}$). Those animats had only two hidden units

305 instead of four but the original type of motors with the possibility of evolving feedback loops.
306 Again, the possible integration of motor units allows one to utilize information about past
307 movements directly (e.g., like the sensation of one's legs). Finally, we included a condition
308 with larger MBs with eight hidden units and motor feedback ($G_{bigbrain}$).

309 We observed that fitness and reliability across group sizes in the original environment
310 decreased for animats with fewer computational units (see Fig 5 and Fig 6). However, while
311 animats in $G_{smallbrain}$ still evolved to reasonably high fitness and reliability, $G_{!feedback}$ was
312 lacking in both. This indicates that motor feedback facilitates evolution in our task
313 environment. One behavioral difference between these two conditions was the reduced
314 movement in the animats of $G_{smallbrain}$ (see Fig 6(c)). Furthermore, the state transition analysis
315 shows that the motor units of animats in $G_{smallbrain}$ tend to change their behavior more often,
316 while animats in $G_{!feedback}$ stay in the same state more often (see Table 4). Notably, $G_{!feedback}$
317 and, particularly, $G_{smallbrain}$ performed better than $G_{0.50}$ given large changes in the wall
318 arrangement.

319 **Fig 5. Fitness evolution and distribution of the final evolved fitness. (a)** Less capacity for memory
320 and internal computations impairs fitness evolution. Despite their similar capacity for memory,
321 $G_{smallbrain}$ evolved higher fitness than $G_{!feedback}$. **(b)** Ceiling outlier suggest that animats in $G_{!feedback}$
322 are generally capable of performing as well as average animats in $G_{smallbrain}$ but that this is less likely.
323 The performance of $G_{bigbrain}$ is comparable to $G_{0.50}$ with more distributed outcomes.

324 **Fig 6. Reliability tests. (a)** $G_{smallbrain}$ is more reliable than $G_{!feedback}$. Considering $G_{bigbrain}$, animats in
325 this group are overall comparable to the baseline condition $G_{0.50}$, but show worse performance in the
326 *Blocked* test condition and some of the modified environments for larger group sizes. **(b)** Reliability
327 R correlates with EF for all setups. The lower reliability of $G_{smallbrain}$ and $G_{!feedback}$ compared to
328 baseline can thus be explained by their already lower evolved fitness values. Note, however, that
329 $G_{smallbrain}$ and $G_{!feedback}$ perform better than $G_{0.50}$ across group sizes in the 4 (*Messy*) *Rooms* test
330 conditions (see **(a)**). **(c)** For larger group sizes, $G_{smallbrain}$ remains static more often than $G_{!feedback}$.

331 **Table 4:** Absolute difference between the state transition probability of $G_{smallbrain}$ and $G_{feedback}$. The
 332 first digit describes whether anything (wall or other animat) is sensed (1) or not sensed (0) and the
 333 second digit describes whether the animat moved/turned (1) or did not move/turn (0). Most notably,
 334 animats in $G_{smallbrain}$ switched more often between sensing and moving than animats in $G_{Feedback}$
 335 (“01→10”, “10→01”, but “11→11”).

$t \backslash SM$	00	01	10	11
00	0.0000	0.0001	0.0000	0.0000
01	0.0000	-0.0167 ¹	0.0237	-0.0046
10	0.0000	0.0194	0.0011	0.0029
11	0.0001	-0.0004	-0.0015	-0.0241

336 ¹ Minus values indicate that the transition is more frequent in $G_{feedback}$, while positive values indicate the
 337 opposite.

338 By contrast, more hidden units ($G_{bigbrain}$) do not improve average fitness or reliability in
 339 any of the tested conditions (see Fig 5 and Fig 6). While $G_{bigbrain}$ overall seems very similar
 340 to the baseline setup $G_{0.50}$, differences can be observed in the *Blocked* and *Small Gate* test
 341 conditions (see Fig 6(a)). In principle, more computational units allow for better
 342 performance. However, the larger space of possible solutions may also impede fitness
 343 evolution (note the larger variance for $G_{bigbrain}$ compared to $G_{0.50}$ in Fig 5(b) and Fig 6(b)).

344 Considering brain complexity, the evolutionary setups with smaller MBs ($G_{smallbrain}$ and
 345 $G_{feedback}$) have significantly lower Φ^{Max} and fewer concepts than the baseline condition
 346 ($G_{0.50}$). Between those two conditions, $G_{smallbrain}$ shows significantly higher Φ^{Max} and more
 347 concepts as compared to $G_{feedback}$ (see Fig 7). This correlates with the larger evolved fitness
 348 values of $G_{smallbrain}$ in Fig 5 and its associated higher reliability in Fig 6. Note that calculating
 349 Φ^{Max} and the corresponding number of concepts was not possible for $G_{bigbrain}$ since
 350 exhaustive evaluations across many systems and states are not currently feasible when using
 351 the *pyphi* software package to compute measures of integrated information theory for
 352 networks of that size (>10 units) [23].

353 **Fig 7. Distribution of brain complexity measures.** Compared to the baseline, the smaller MBs
 354 ($G_{smallbrain}$ and $G_{feedback}$) have lower Φ^{Max} and fewer corresponding concepts. Animats in $G_{smallbrain}$
 355 show higher Φ^{Max} and have more corresponding concepts compared to $G_{feedback}$ animats, many of

356 which have $\Phi^{Max} = 0$. Due to computational reasons, the brain complexity of $G_{bigbrain}$ could not be
357 calculated (see text).

358 **Varying interaction conditions: Evolution of beneficial interaction**

359 In our evolution simulations, the fitness function used for selection depended on the
360 average task fitness of all animats in the group. Moreover, individuals received penalties for
361 colliding with other group members. Since it is hardly possible to directly observe
362 cooperative interactions, we used a third set of simulations to manipulate aspects of the
363 fitness function and physical interaction between animats to identify to what extent these
364 features influence both the fitness and the reliability. For this purpose, we considered four
365 different evolutionary setups besides the baseline setup $G_{0.50}$: G_{single} (same as above), $G_{!penalty}$,
366 $G_{blocked}$, and $G_{blocked!/penalty}$ (see Table 1 for a detailed description). G_{random} is also included in
367 the figures for comparison.

368 Among the novel setups, only animats in $G_{blocked}$ were subject to the collision penalty
369 during evolution, whereas later, during the *Original* reliability tests, all conditions were
370 subject to a penalty. Not being able to share the same position (as in $G_{blocked}$) hardly
371 influenced the final fitness, reliability, or behavior of the evolved animats (Fig 8 and Fig 9,
372 compared to the baseline condition). $G_{!penalty}$, where reacting to other animats had no direct
373 effect on the fitness evolution, showed very similar fitness evolution, reliability curves, and
374 behavior to G_{single} . Considering the reliability tests in Fig 9(a), the top row shows the
375 reliability across group sizes in the *Original* environment, and under varying interaction
376 conditions: *No Penalty*, *Blocked*, and both *Blocked and no Penalty* (from left to right). In the
377 bottom row of Fig 9(a), animats are evaluated under the same interaction rules as they
378 evolved in while only facing a modified environment.

379 **Fig 8. Fitness Evolution and distribution of the final evolved fitness.** The animats in conditions
380 without a penalty ($G_{blocked!/penalty}$ and $G_{!penalty}$) evolved to relatively high fitness levels. In particular,

381 $G_{!penalty}$ evolved like G_{single} , since animats in both conditions were not impacted at all by other
382 animats. Similarly, $G_{blocked}$ seemed equivalent to the baseline setup $G_{0.50}$, while $G_{blocked!/penalty}$ evolved
383 to slightly higher fitness values, comparable to G_{random} .

384 **Fig 9. Reliability Tests.** (a) There was a significant difference between animats in $G_{blocked!/penalty}$ and
385 animats in $G_{!penalty}$. Being blocked was essential for retaining some reliability if no penalty was given.
386 (b) $G_{!penalty}$ showed similar reliability as G_{single} , whereas $G_{blocked}$ showed similar reliability as $G_{0.50}$.
387 (c) These similarities were also reflected in the animats' behavior. The behavior of animats in
388 $G_{blocked!/penalty}$ was more reactive to changing group size than $G_{!penalty}$.

389 Comparing the reliability tests of $G_{blocked!/penalty}$, $G_{blocked}$ and $G_{!penalty}$ (Fig 9), we observed
390 significant differences between the setups, which let us assume that there is implicit
391 cooperation. In this context, we want to highlight that $G_{!penalty}$ performs relatively poor for
392 larger group sizes in the environment designs with large modifications (in *4 (Messy) Rooms*)
393 as compared to the other setups. This is an indicator for the evolution of beneficial
394 interactions between group members in evolutionary setups with a collision penalty and/or
395 blocking. The decline in task fitness of $G_{blocked!/penalty}$ for higher group sizes under test
396 conditions with a collision penalty showed that these animats did not avoid physical
397 interactions with their group members, while $G_{blocked}$ animats were generally comparable to
398 $G_{0.50}$. However, even $G_{blocked!/penalty}$ animats had an advantage compared to $G_{!penalty}$ in the *4*
399 (*Messy*) *Rooms* environment, which may be due to some implicit form of cooperative
400 behavior.

401 Considering the brain complexity of animats in $G_{blocked}$ and $G_{blocked!/penalty}$, we can report
402 similar values to $G_{0.50}$ (see Fig 10). Whether animats received a penalty for crossing each
403 other, or whether crossing was prohibited to start with, did not significantly affect their
404 evolved fitness, reliability, behavior, or brain complexity. Likewise, the brain complexity
405 measures for $G_{!penalty}$ were comparable to those of G_{single} , in line with the behavioral results
406 above.

407 **Fig 10. Distribution of brain complexity measures.** In evolutionary setups where crossing each
408 other was not possible ($G_{blocked}$ and $G_{blocked/penalty}$), the brain complexity was comparable to the
409 complexity of $G_{0.50}$. By contrast, animats in setups where the reaction to fellow animats had no
410 reasonable effect on their performance (G_{single} and $G_{penalty}$) showed lower brain complexity. Still,
411 there was high variance in the data of brain complexity.

412 **Varying sensor configuration: Sensory capacity influences reliability and intrinsic** 413 **complexity**

414 Finally, we manipulated the ability of dealing with the task (task difficulty) by changing
415 the sensor configuration of the animats. In addition to the baseline architecture, we designed
416 animats with sensors on three sides G_{3sides} (front, left and right), without an agent sensor
417 $G_{!agent}$ and with a universal sensor $G_{w=a}$ (sensing wall and agent as indiscriminate obstacles).
418 Fig 11 reveals that it is necessary to have the ability to sense nearby animats, and be able to
419 differentiate between walls and animats, in order to achieve reasonable fitness values.
420 Generally, it was an advantage to be equipped with sensors on more sides for both high task
421 fitness and high reliability.

422 **Fig 11. Fitness Evolution and distribution of the final evolved fitness.** The average evolved fitness
423 showed that animats in evolutionary setups without specific sensors for other animas ($G_{!agent}$ and
424 $G_{w=a}$) achieved no reasonable fitness. By contrast, animats in G_{3sides} outperformed $G_{0.50}$, and G_{random} ,
425 but also had more outliers with lower fitness and performed worse than the baseline condition in
426 early generations (up to ~10k generations).

427 Regarding reliability, we would first like to highlight animats in the G_{3sides} condition.
428 They consistently outperformed the animats in other groups except in two test conditions:
429 *Blocked* and *Noisy Corners* (see Fig 12). This shows that animats which are equipped with
430 more sensors do have an advantage on average, but they may also perform worse than animats
431 with fewer sensors under some circumstances. The sensory signals in these specific

432 environments might have been too different from the information patterns the animats
433 evolved in and were thus specialized for.

434 **Fig 12. Reliability Tests. (a-b)** The G_{3sides} condition was the most reliable in most test conditions,
435 except in *Blocked* and *Noisy Corners*. In terms of reliability, sensing everything ($G_{w=a}$) with one
436 sensor is still better than only sensing the walls due to a missing animat sensor ($G_{!agent}$). **(c)** Setups
437 with few sensors evolved no general behavior (high variance of movement between the 30 different
438 evolutions, shaded area). The G_{3sides} setup becomes more reactive as soon as the animat density starts
439 to rise.

440 **Fig 13. Distribution of brain complexity measures.** Animats in the G_{3sides} condition showed the
441 lowest brain complexity of all setups despite having the highest evolved fitness and reliability. By
442 contrast, animats with limited sensor information ($G_{!agent}$ and $G_{w=a}$) had lower than baseline
443 complexity values, but also low evolved fitness (EF , see Fig 11).

444 Opposite behaviors can be observed for the animats in $G_{w=a}$ and $G_{!agent}$. In this case,
445 animats were not evolving to reasonable fitness values. Nevertheless, we could observe
446 differences between the two conditions from their reliability values. While $G_{w=a}$ animats had
447 only one sensor which does not discriminate between the wall and other animats, $G_{!agent}$ was
448 missing the animat sensor completely. $G_{!agent}$ showed better task fitness than $G_{w=a}$ in test
449 conditions with small group sizes and without a penalty. Considering the evolved behavior,
450 $G_{w=a}$ animats (Fig 12(c)) were not reactive to other animats, which suggests that they did not
451 evolve the capacity to differentiate between the animats and the walls internally, e.g., through
452 memory.

453 Analyzing the brain complexity showed that animats equipped with fewer, but also with
454 more sensors than in the baseline setup $G_{0.50}$ evolved MBs with lower complexity (see Fig
455 13), albeit for different reasons. Based on the very low evolved fitness for $G_{w=a}$ and $G_{!agent}$
456 (see Fig 11) we can conclude that their MBs did not develop the necessary structure and
457 mechanisms to solve the task, as reflected by their low brain complexity. By contrast, animats

458 in G_{3sides} achieved high performance and reliability, but did not evolve any integrated
459 information ($\Phi^{Max} = 0$) in many cases. This observation was in line with previous findings
460 on the relation between sensory capacity and internal complexity [19] and suggested that
461 high brain complexity in cognitive systems depends on a need for internal memory and
462 computation, which may decrease if an animat is equipped with more sensors. Please refer
463 to the next section for a general discussion about the relationship between task performance,
464 reliability, and brain complexity.

465 **Discussion**

466 The evolution of cooperative multi-agent systems might be the next frontier in the
467 context of evolving artificial agents, in which context not much is yet known about conditions
468 that give rise to cooperative behavior and the complex inter-dependencies between individual
469 and group goals [24]. For example, there might be many factors that influence whether the
470 individuals either bow to the group or act by egoistic rules [25]. In this study, we used animats
471 equipped with MBs (introduced by Edlund et al. [21]) to study how group performance and
472 its reliability under modified conditions depended on the individual, interactions between
473 individuals, as well as specific features of the MBs' evolution.

474 **Prior work investigating group evolution**

475 Earlier research that implemented groups of MBs concentrated on predator-prey
476 environments and showed that animats can (co-)evolve swarm behaviors [26–28]. The
477 animat design in this work was generally based on a design in Marstaller et al. [15], who
478 evolved individual MBs with the goal of solving perceptual-categorization tasks. Another
479 method of simulating swarm behavior is neuro-evolution, i.e., the evolution of *artificial*
480 *neural networks (ANN)* [29–31]. As in Olson et al. [27], these neuro-evolution experiments
481 produced agents which evolve in a swarm to solve a predator-prey task.

482 Other researchers have investigated the effect of group size in the evolution of groups of
483 simulated agents beyond predator-prey scenarios in a more general context. They find that
484 the behavior of the group of agents and the individual agent is dependent on the group size
485 [32,33]. In another study which changed the group size during evolution, the authors show
486 that it can be easier for smaller groups than larger ones to organize themselves [5].

487 The effect of changing swarm sizes has also been investigated in the context of natural
488 biological systems: Brown [25] examined which factors are decisive for the individual to
489 either join a swarm or behave egoistically. The study focused on experimenting with
490 environmental qualities and swarm size. Brown defined *optimal swarm size* as the best trade-
491 off between the advantage of balancing costs between individuals in the swarm and the
492 disadvantage of sharing the resources (energy/food) with the whole swarm. In an earlier
493 study, Pacala et al. [4] report that swarm size constrains information transfer and task
494 allocation. They argue that the information exchange varies and the task allocation changes,
495 depending on the swarm size of ant-colonies. Pacala et al. [4] also argue that swarm behavior
496 is the product of social interaction, individual interaction, and the interaction with the given
497 environment. In a more recent work [34], we found arguments that swarm behavior arises if
498 there is sufficient density within the swarm.

499 **Factors that impact task performance and reliability**

500 In line with the variety of dependencies identified in these earlier studies, our simulation
501 results suggest that group performance and reliability under modified conditions are complex
502 multidimensional phenomena. Our work is illustrative, as it shows that there is high
503 complexity even in the simplified experimental setting of small artificial organisms evolving
504 within a particular evolutionary setup which is completely controlled by the experimenter.
505 Nevertheless, by creating a variety of environments and animats, we were able to identify
506 several factors that influence fitness evolution and post-evolutionary reliability.

507 Generally, task difficulty (the ability to evolve high fitness in a given task environment)
508 depends on the complexity of the environment, but also on the animats' architecture (see also
509 [19]). In the specific evolutionary setup investigated here, evolved fitness negatively
510 correlated with group size as a result of the imposed penalty for collisions (see Fig 2). On the
511 other hand, animats evolved in fixed, intermediate group sizes are most reliable to changes
512 in group size, and, in fact, comparable to animats evolved for reliability that experienced
513 random group sizes during evolution (Fig 3(b)). Yet animats evolved in large groups
514 performed slightly better in modified environments (Fig 3(a), bottom row). A similar trade-
515 off can be observed for different animat architectures: animats with less capacity for memory
516 ($G_{smallbrain}$ and $G_{feedback}$) evolved to lower fitness levels than the baseline condition ($G_{0.50}$)
517 (Fig 5), and were less reliable under changes in group sizes but still showed better
518 performance in some of the modified environments (Fig 6(a)). More hidden units ($G_{bigbrain}$)
519 did not provide further advantages compared to $G_{0.50}$. Finally, more sensors (G_{3sides}) proved
520 advantageous for both evolved fitness and reliability under almost all modified test
521 conditions. However, even G_{3sides} performed worse than the baseline in one of the modified
522 environments (*Noisy Corners*). Within most specific environmental setups, reliability to
523 changes in group size was, moreover, correlated with evolved fitness (Figs 3/6/9/12 (b), right
524 panel).

525 Overall, we found that the right balance is essential: If the environmental design is
526 balanced to the animats' architecture (having the right sensor setup, memory capacity, and
527 motor setup), animats evolved consistent reliability, even if it was not specifically trained for.
528 In other words, animats that were well-equipped for dealing with their original task
529 environment (and thus achieved high evolved fitness) were generally also able to remain
530 reliable given small modifications to task conditions. However, evolutionary setups that seem
531 less adapted (lower evolved fitness) overall may still have advantages under some conditions.

532 **Interactions between individuals in the group**

533 In this study, we did not explicitly implement any form of direct communication between
534 animats. Nevertheless, through triangulation, we can partly answer whether the evolutionary
535 setup we employed here may have led to the evolution of implicit cooperation between group
536 members. To that end, we have shown that it was necessary for animats to perceive their
537 fellow group members, and that they use this information to achieve reasonable evolved
538 fitness and reliability (Fig 11 and Fig 12). Moreover, animats evolved in large groups showed
539 an advantage across group sizes in modified environments (Fig 3(a), bottom), while animats
540 that evolved without a collision penalty (G_{penalty}) performed worse in some of the modified
541 environments, even if tested without a penalty (Fig 9(a), 4 (*Messy Rooms*)).

542 Hypothetically, this type of implicit interaction between animats is less related to *verbal*
543 *communication*, but it may relate more to communication through behavior (e.g., like bees
544 performing their dance). As we know from previous studies, swarm behavior in nature can
545 also be the result of simple reactions to local neighbors [3,35]. We argue that animats are
546 interdependent in this way, even if there is no explicit information exchange between them.
547 The observed instances of cooperative behavior can thus be viewed as an emergent
548 phenomenon of the evolutionary process.

549 **Relation between brain complexity, task performance, and reliability**

550 Previous studies applying measures of integrated information to adaptive animats
551 equipped with MBs [19,21,36] have observed that Φ^{Max} and related measures on average
552 increase over the course of evolution, which correlates with increasing task performance (see
553 Table S6 in Supporting Information S2). Moreover, as demonstrated in [19], this increase
554 depends on the complexity of the task environment relative to the animats' sensor capacity:
555 MBs that evolved in task environments which required more memory and internal
556 computation developed, on average, higher Φ^{Max} values and a higher number of concepts.

557 For the evolutionary setups with the standard animat architecture as in $G_{0.50}$, we found
558 the highest values of Φ^{Max} and $\#Concepts(\Phi^{Max})$ for medium group sizes $G_{0.50}$, and $G_{blocked}$,
559 and for G_{random} . These setups were also among the most reliable across group sizes (see also
560 [17] for similar results using a simplified measure of brain complexity). By contrast,
561 significantly lower Φ^{Max} values were found for G_{single} and $G_{!penalty}$, the two setups in which
562 task fitness during evolution did not depend on interactions with other animats. As argued
563 above, G_{single} and $G_{!penalty}$ thus effectively evolved within a simpler task environment than
564 $G_{0.50}$, $G_{blocked}$, and G_{random} , which explains their lower Φ^{Max} .

565 Compared to $G_{0.50}$, evolutionary setups with altered animat architectures showed
566 consistently lower values of Φ^{Max} and $\#Concepts(\Phi^{Max})$. Limiting the animats' sensor
567 capacity ($G_{!agent}$ and $G_{w=a}$) or the number of available memory units ($G_{smallbrain}$ and $G_{!feedback}$)
568 interfered with their capacity for successful evolution in the spatial navigation task. Their
569 lower performance was thus accompanied by less developed MBs with lower Φ^{Max} and fewer
570 concepts. Given more time to evolve (more generations), both their performance and their
571 brain complexity might still increase. By contrast, more sensors allowed for better
572 performance based on high amounts of external information, which effectively decreased the
573 need for internal complexity (memory and computations) and thus may also lead to low Φ^{Max} ,
574 as observed here for G_{3sides} .

575 In theory, high fitness in any given environment could be achieved without information
576 integration (e.g., by a system with a large feed-forward architecture [18]), and information
577 integration can be high even if there is no reasonable fitness, which partially explains the
578 large variance in the brain complexity measures (see, e.g., outliers for $G_{!agent}$ in Fig 13)
579 However, given a certain requirement for memory and context sensitivity, constraints in the
580 number of sensors and hidden elements may give rise to an empirical lower boundary on the
581 amount of integrated information necessary to perform a given task [19,21,36,37].

582 In summary, for a given MB architecture, higher brain complexity seems to be related to
583 better performance and reliability. However, future work should explore under which
584 environmental conditions additional sensors, or more internal units, become more
585 advantageous for the evolution of higher task performance and reliability.

586 **Limitations**

587 Our work modeled one particular, small-scale scenario. Future work should consider
588 other task environments which may strengthen the generality of our results. Moreover, further
589 evolution or training scenarios for artificial organisms should be considered as well.

590 While the measures that we employed to assess the complexity of the evolved MBs are
591 theoretically motivated [18], they are also computationally very complex. This made it
592 difficult to evaluate a larger number of evolution simulations in order to achieve better
593 statistical power. This is why alternative, approximate measures should be considered, too.
594 For instance, the *largest strongly connected component* (and other graph metrics) can be used
595 as a proxy for system integration and thus brain complexity [17]. Efficient approximations
596 would also enable investigation into how brain complexity develops across generations.
597 Moreover, Φ^{Max} , and the associated number of concepts, are causal measures that assess the
598 degree to which the mechanisms within a MB are differentiated and integrated. Future work
599 should also consider and explore alternative informational or dynamical measures [e.g., 38–
600 40]. In this study, we concentrated on the reliability tests, so the brain complexity analysis
601 was not the subject of more in-depth investigation.

602 **Conclusion**

603 It is challenging to remain reliable in a dynamic and volatile world while also trying to
604 succeed in a given task. So, investigating the characteristics of this reliability might help to
605 develop implications and strategies for improving reliability. We showed that reliability is a

606 complex concept to investigate, especially when considering not only individuals but an
607 organized group. Yet we were able to isolate essential influencing factors to better understand
608 the positive and negative effects of changing group size, environment design, and individual
609 cognitive ability on task reliability. This research asserts that task efficiency and effectiveness
610 is not the only goal; task reliability is also worth striving for. We have also offered a
611 computational approach for investigating this concept.

612 **Materials and Methods**

613 We used an EA to generate simulated animats evolving in groups, and defined and tested
614 various animat architectures and evolutionary environments to evolve animats having
615 heterogeneous behavior, fitness, and reliability. Afterwards, we conducted post-evolutionary
616 tests to assess the reliability of the different evolutionary setups. This section explains the
617 animat designs, the environment, the evolutionary simulations, and the experiment setup. We
618 used *MABE (Modular Agent-Based Evolver)* [41] as a computational evolution framework
619 with the same parameters as in previous work [17] (see Table S7 in Supporting Information).

620 As we state in the introduction, we studied the changes in behavior and task performance
621 of evolved animats while manipulating environmental and cognitive conditions, which also
622 changed the ability to achieve the goal of the task (task difficulty). The idea was that the
623 individual animat had to solve a two-dimensional spatial-navigation task, thus forcing
624 individuals to react to other animats in order to reach a high fitness value. This task was a
625 redesign by Fischer et al. [17] of a task environment initially developed by Koenig et al. [20].
626 An animat can usually differentiate between static (borders and walls) and dynamic objects
627 (animats) in the environment through two distinct sensors. This design allowed for the
628 evolution of social behavior based on passive interactions between animats (we observed,
629 e.g., “waiting”, or “following” behavior).

630 **Animats Architecture**

631 The EA evolves animats with MBs, which contain a set of discrete, binary computational
632 units (“neurons”). Each unit has its own update rules receiving inputs from and sending their
633 output to other units. In this study, the decision system (the connectivity between units and
634 their update-rules) was implemented by *Hidden Markov Gates (HMGs)*. The HMGs connect
635 the nodes of the MB indirectly. Fig 14 visualizes a simple example, in which an HMG is
636 connected to four units. The decision system inside an HMG can be diverse. In this research,
637 we evolved discrete lookup tables. The lookup tables translate the states of the connected
638 input units at t to the new states of connected output units at $t+1$. The motor or memory units
639 can represent the output units of the HMG. In this study, the EA evolved genomes with a
640 string of natural numbers. The individual numbers encoded the HMGs: the number of HMGs,
641 the lookup tables, the connected input units, and the connected output units. The EA mutated
642 the genomes in each generation. Each locus in the genome mutated with a certain probability.
643 In addition, larger sections could be deleted or added to the genome [21,42] (again, all
644 parameters are listed in Table S7 within the Supporting Information).

645 **Fig 14. Example of an MB.** An MB [21] has three components: (1) Units with a binary states (“1”-
646 “4”), (2) HMGs and (3) the connections between the binary units and the HMGs. The connections
647 between the units can be derived from the connections to the HMGs. HMGs contain the mechanism,
648 e.g., a probabilistic lookup table, to transform the brain state of units at t to the state at $t+1$.

649 All units in the animat’s MB have binary states, either 1 or 0 , e.g., a sensor turns 1 if an
650 obstacle is detected and a motor switches to 1 if it is active. Two motors provide the ability
651 to turn 90 degrees left or right, and to move forward (if both motors are in state 1). Since the
652 units within a MB can be interconnected in a recurrent manner, they have the potential to
653 create internal memory. We evolved animats with five different animat designs. Fig 15 gives
654 a schematic overview of all animat designs. In addition to the baseline cognitive architecture,
655 which was introduced already in [17], further deviations were designed to investigate the

656 influence of different cognitive setups on the resulting evolved behavior, task performance,
657 and reliability. The sensors had a detection range of one unit. Typically, the motor units could
658 also feedback to the hidden and motor units, thus acting as additional brain capacity, since
659 knowledge about previous motor states is directly available for computing the next state.
660 Additionally, we designed an animat without motor feedback ($G_{!feedback}$).

661 **Fig 15. Schematic architecture of the five different animat designs.** The animats have two motor
662 units (grey triangles), four hidden units (dark grey circles) and one to six sensor units (black/red
663 shapes). **(a)** Baseline design as in [17]. **(b)** Animat with sensors on three sides. There is an animat
664 sensor and a wall sensor on each side. **(c)** Animat without feedback motors (motors cannot be part of
665 the memory network). **(d)** Animat with a single sensor unit, measuring wall and animat
666 simultaneously. **(e)** Animat without an animat sensor. Note that the architectures depict the maximal
667 amount of units available. Whether any given unit is actually used depends on the evolved
668 connectivity and logic function. Animats are initialized without connections between units.

669 Design of the 2D Environment

670 All experiments simulated a two-dimensional environment. The world has 32×32 units
671 (see Fig 16). All animats started on one of 72 predefined, uniformly distributed, starting
672 positions. The selection for the starting position, as well as an animat's initial orientation,
673 was random. The original environment (see Fig 16(a)) had two rooms, which are connected
674 by a gate. The animats' goal was to travel between the two rooms in order to achieve a high
675 fitness value. This design was adapted from the work of Koenig et al. [20]. As an additional
676 dimension for evaluating reliability under environmental change, we tested all evolved MBs
677 (the final generation) in additional environment designs (see Fig 16(b-e)).

678 **Fig 16. Environmental design.** **(a)** The two-dimensional environment is based on a discrete grid
679 architecture and contains two rooms. Animats draw a random starting position. Their orientation can
680 be up, down, left, and right and is also randomly selected at initiation. **(b-e)** Four additional rooms
681 were used to test the reliability of the animats. Red blocks mark the changes/additions in the room

682 and represent walls. In (d), all four gates count as possible rewards. In (e), only gates on the vertical
683 mid-line provide rewards.

684 We chose MBs as a simplified model of an artificial brain, since the basic idea of an MB is
685 to emulate the recurrent connectivity structure found in real neural networks in a simple
686 manner, while being complex enough to represent a cognitive system [15]. Furthermore, a
687 recent study showed that MBs can be very compatible against variations of *artificial neural*
688 *networks* and even showed higher performance in general [16]. Nevertheless, it would, in
689 principle, also be possible to use a finite state machine [20], or artificial neural networks [30]
690 to solve these kinds of tasks.

691 Experiment Design

692 We selected $G_{0.50}$ to be the baseline setup for evolution, to which we compared all other
693 evolutionary setups. This was because $G_{0.50}$ showed the highest reliability across group sizes.
694 In sum, we came up with **15** different setups for the evolution of the animats. Using the
695 MABE framework, we simulated each evolutionary setup **30** times. In each of these **30**
696 evolutions, the EA had **10,000** generations to converge on the final solution. **100** genomes
697 were mutated and evaluated in each generation. Each of these evaluations was repeated **30**
698 times with different starting positions, orientation, and selection order (for the serial
699 processing of the animats' movement). After a genome was tested 30 times, it received a
700 fitness score, which was computed based on the mean across the task performance of 30
701 single animats, with one being picked randomly from each of the **30** random test runs. In
702 addition, in setup G_{random} the group size varied for each of the **30** tests (drawn randomly from
703 $72*[1, 0.95, 0.9, \dots, 0.1, 0.05, 0.01389]$).

704 **The Simulated Life**

705 The fitness function that determines the probability of a genome being reproduced
706 depends on two factors. First, animats have to travel as often as possible through the gate
707 (change the room) (see Fig 16). Second, the animats need to avoid colliding with each other.
708 Fischer et al. [17] showed the formal definitions of the fitness function as a weighted sum of
709 the penalty for collision and the reward for crossing the gate. The weight of the reward (factor
710 **1.0**) is higher than the weight in the case of a penalty (factor **0.075**). These weights need to
711 be chosen carefully. If the penalty is too low or the reward is too high, animats will keep
712 moving from one room to the other through the gate (herding effect) and ignore the penalty.
713 On the other hand, given a high penalty and low reward, animats will evolve hardly any
714 movement. To further reduce the herding effect around the gate, there is a refractory period
715 of **100** timesteps after receiving a reward before an animat can receive another reward. Since
716 each trial has a duration of **500** timesteps, any one animat can receive a total fitness score of
717 at most **4** [17].

718 To further raise the task difficulty and to investigate the coordination and cooperation of
719 animats in groups, we let animats co-exist in the same environment (in contrast to previous
720 studies in this scope [15,18,21]). Currently, we have not implemented co-evolution and have
721 only evaluated a genome by generating animats as identical clones (they have the same MB).
722 There was no active knowledge exchange (“communication”) between animats in this study.
723 Through the architecture of the animats, they have to develop the ability to distinguish which
724 kind of sensory input to use for decision making. Sensors can only sense one position in front
725 of (or on the side of) the animat and differentiate between static objects (walls) and dynamic
726 objects (fellow animats), except for $G_{w=a}$.

727 Compared to the baseline setup, we included further control conditions in which animats
728 did not receive the collision penalty and/or were not able to overlap. Those changes in the

729 fitness function represented environmental rules which influenced the task difficulty. As a
730 result, we were able to test dependencies between the evolution environment and the
731 evolution of reliability.

732 **Post-Evolutional Evaluation**

733 **Reliability tests.** The reliability tests were designed as follows: First, we selected the **30**
734 genomes of generation **10,000 (10k)** for each of the **15** conditions. Second, each genome was
735 tested across **21** conditions varying in group size. To this end, we created groups of animat
736 clones of the respective test group size for each of the **30*15** genomes. Test group sizes were
737 uniformly distributed between **1** and **72**. The interval of the relative distribution is **[0.0139,**
738 **0.05, 0.1, ..., 0.9, 0.95, 1.0]**. A single animat is obviously not a group, but we treat it as one
739 in order to simplify notation.

740 In addition to the reliability tests across varying group sizes in the baseline task design
741 (*Original*), we created four modified test environments, as shown in Fig 16 (*Noisy Corners,*
742 *Small Gate, 4 Rooms, 4 Messy Rooms*). Moreover, we included three additional test
743 conditions in which we varied the interaction properties of the animats (*No Penalty, Blocked,*
744 *Blocked and no penalty*). Finally, we tested each of the **30*15*21** different configurations in
745 each of the eight test environments.

746 For the statistical analysis and the main reliability evaluations, we defined a reliability
747 measure across group sizes in the *Original* environment design: $R = (TF)_{GS}$. The modified
748 test environments represented four independent samples of possible environmental
749 modifications and were only evaluated on their own for this reason. The results of the
750 remaining three test conditions with varying interaction properties mainly served to highlight
751 differences between the evolutionary setups, rather than testing reliability per se.

752 **Brain complexity.** To evaluate the complexity of the evolved MBs, we employed two
753 complimentary measures provided by integrated information theory (IIT) [18,43], Φ^{Max} and
754 the associated number concepts ($\#Concepts(\Phi^{Max})$). A major advantage of the measures
755 developed within the IIT framework is that we can quantify the internal mechanisms (causal
756 relations) of animats and their interactions (e.g., [44,45]), which let us construct premises on
757 how the cognitive processes work. The core of IIT's measures is an information theoretic,
758 and probabilistic graph analysis [18] based on the state-to-state transition probabilities of the
759 units, i.e., their update functions. Please refer to [18,19] for details on the evaluation. All
760 calculations were conducted using the IIT Python package *pyphi* [23], which we used in our
761 work to calculate Φ^{Max} and the corresponding number of concepts. Φ^{Max} represents the
762 highest possible integrated information the system can achieve across all its subsets, which
763 we used as an indicator for brain complexity. A concept is a set of physical mechanisms (e.g.,
764 neurons) that create integrated information [18]. Since the employed measures are state-
765 dependent, we evaluated Φ^{Max} and the number of concepts for every state a MB experienced
766 during a lifetime (one trial) and selected the maximum value over all states as in [19]. Fig S1
767 (Supporting Information S2) shows by way of example that it is essential for high Φ^{Max} in a
768 system that many elements be integrated, meaning also maintaining feedback loops within
769 the system. In this study, we only considered the brain complexity of the final generation
770 (**10k**) due to the computational complexity of calculations using *pyphi*.

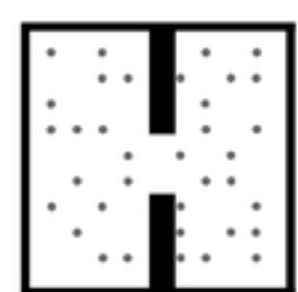
771 **Statistics.** The evolved fitness values, the reliability R , and the IIT brain complexity
772 measures were statistically evaluated across all evolutionary setups using a Kruskal-Wallis
773 test, which showed a significant difference of the observed statistics between all groups taken
774 together. Further, we used the Mann-Whitney-U test to evaluate the difference between pairs
775 of evolutionary setups. Section S3 in the Supporting Information lists all statistical tests that
776 are a subject of discussion in the results and discussion section.

777 **References**

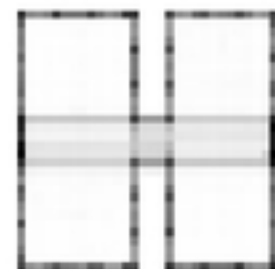
- 778 1. Spearman C. “General Intelligence,” Objectively Determined and Measured. *Am J Psychol.* 1904;15: 201–292.
- 779 2. Gardner H. The theory of multiple intelligences. *Ann Dyslexia.* 1987;37: 19–35. doi:10.1007/BF02648057
- 780 3. Garnier S, Gautrais J, Theraulaz G. The biological principles of swarm intelligence. *Swarm Intell.* 2007;1: 3–31.
- 781 doi:10.1007/s11721-007-0004-y
- 782 4. Pacala SW, Gordon DM, Godfray HCJ. Effects of social group size on information transfer and task allocation.
- 783 *Evol Ecol.* 1996;10: 127–165. doi:10.1007/BF01241782
- 784 5. Dorigo M, Trianni V, Şahin E, Groß R, Labella TH, Baldassarre G, et al. Evolving Self-Organizing Behaviors for
- 785 a Swarm-Bot. *Auton Robots.* 2004;17: 223–245. doi:10.1023/B:AURO.0000033973.24945.f3
- 786 6. Weick KE, Sutcliffe KM, Obstfeld D. Organizing for High Reliability: Process of Collective Mindfulness. *Crisis*
- 787 *Management.* 2008. doi:10.1177/0020764009106599
- 788 7. Weick KE, Roberts KH. Collective Mind in Organizations: Heedful Interrelating on Flight Decks. *Adm Sci Q.*
- 789 1993;38: 357. doi:10.2307/2393372
- 790 8. Oliver N, Senturk M, Calvard TS, Potocnik K, Tomasella M. Collective Mindfulness, Resilience and Team
- 791 Performance. *Acad Manag Proc.* 2017;2017: 12905. doi:10.5465/AMBPP.2017.12905abstract
- 792 9. Pinter-Wollman N, Penn A, Theraulaz G, Fiore SM. Interdisciplinary approaches for uncovering the impacts of
- 793 architecture on collective behaviour. *Philos Trans R Soc B Biol Sci.* 2018;373. doi:10.1098/rstb.2017.0232
- 794 10. Engel D, Malone TW. Integrated information as a metric for group interaction. *Dovrolis C, editor. PLoS One.*
- 795 2018;13. doi:10.1371/journal.pone.0205335
- 796 11. List C, Philip Pettit. Group Agency and Supervenience. *South J Philos.* 2006;44: 1–22.
- 797 12. Walsh J, Ungson GR. Organizational Memory. *Acad Manag Rev.* 1991;16: 57–91. doi:10.2307/258607
- 798 13. Nonaka I. A firm as a knowledge-creating entity: a new perspective on the theory of the firm. *Ind Corp Chang.*
- 799 2000;9: 1–20. doi:10.1093/icc/9.1.1
- 800 14. Tsoukas H. The firm as a distributed knowledge system: A constructionist approach. *Strateg Manag J.* 1996;17:
- 801 11–25. doi:10.1002/smj.4250171104
- 802 15. Marstaller L, Hintze A, Adami C. The Evolution of Representation in Simple Cognitive Networks. *Neural*
- 803 *Comput.* 2013;25: 2079–2107. doi:10.1162/NECO_a_00475
- 804 16. Hintze A, Kirkpatrick D, Adami C. The structure of evolved representations across different substrates for
- 805 artificial intelligence. 2018; Available: <http://arxiv.org/abs/1804.01660>
- 806 17. Fischer D, Mostaghim S, Albantakis L. How swarm size during evolution impacts the behavior, generalizability,
- 807 and brain complexity of animats performing a spatial navigation task. *GECCO 2018.* 2018;
- 808 doi:10.1145/3205455.3205646
- 809 18. Oizumi M, Albantakis L, Tononi G. From the Phenomenology to the Mechanisms of Consciousness: Integrated
- 810 Information Theory 3.0. *Sporns O, editor. PLoS Comput Biol.* 2014;10: 1–25. doi:10.1371/journal.pcbi.1003588
- 811 19. Albantakis L, Hintze A, Koch C, Adami C, Tononi G. Evolution of Integrated Causal Structures in Animats
- 812 Exposed to Environments of Increasing Complexity. *Polani D, editor. PLoS Comput Biol.* 2014;10: e1003966.
- 813 doi:10.1371/journal.pcbi.1003966
- 814 20. König L, Mostaghim S, Schmeck H. Decentralized evolution of robotic behavior using finite state machines.
- 815 Hettiarachchi S, editor. *Int J Intell Comput Cybern.* 2009;2: 695–723. doi:10.1108/17563780911005845
- 816 21. Edlund JA, Chaumont N, Hintze A, Koch C, Tononi G, Adami C. Integrated Information Increases with Fitness
- 817 in the Evolution of Animats. *Graham LJ, editor. PLoS Comput Biol.* 2011;7: e1002236.
- 818 doi:10.1371/journal.pcbi.1002236
- 819 22. Marshall W, Gomez-Ramirez J, Tononi G. Integrated information and state differentiation. *Front Psychol.*

- 820 2016;7. doi:10.3389/fpsyg.2016.00926
- 821 23. Mayner WGP, Marshall W, Albantakis L, Findlay G, Marchman R, Tononi G. PyPhi: A toolbox for integrated
822 information theory. Blackwell KT, editor. *PLOS Comput Biol*. 2018;14: e1006343.
823 doi:10.1371/journal.pcbi.1006343
- 824 24. Miikkulainen R, Feasley E, Johnson L, Karpov I, Rajagopalan P, Rawal A, et al. Multiagent Learning through
825 Neuroevolution. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence*
826 *and Lecture Notes in Bioinformatics)*. 2012. pp. 24–46. doi:10.1007/978-3-642-30687-7_2
- 827 25. Brown JL. Optimal group size in territorial animals. *J Theor Biol*. 1982;95: 793–810. doi:10.1016/0022-
828 5193(82)90354-X
- 829 26. Olson RS. Elucidating the Evolutionary Origins of Collective Animal Behavior. PhD Proposal. 2015.
- 830 27. Olson RS, Hintze A, Dyer FC, Knoester DB, Adami C. Predator confusion is sufficient to evolve swarming
831 behavior. *J R Soc Interface*. 2012;10: 20130305. doi:10.1098/rsif.2013.0305
- 832 28. Olson RS, Knoester DB, Adami C. Critical interplay between density-dependent predation and evolution of the
833 selfish herd. *Proceeding fifteenth Annu Conf Genet Evol Comput Conf - GECCO '13*. 2013; 247.
834 doi:10.1145/2463372.2463394
- 835 29. Karpov I V., Johnson LM, Miikkulainen R. Evaluating team behaviors constructed with human-guided machine
836 learning. 2015 IEEE Conference on Computational Intelligence and Games (CIG). IEEE; 2015. pp. 292–298.
837 doi:10.1109/CIG.2015.7317946
- 838 30. Stanley KO, Cornelius R, Miikkulainen R, Silva TD, Gold A. Real-time Learning in the NERO Video Game.
839 *Proc First Artif Intell Interact Digit Entertain Conf*. 2005;2003: 2003–2004.
- 840 31. Stanley KO, Bryant BD, Miikkulainen R. Real-time neuroevolution in the NERO video game. *IEEE Trans Evol*
841 *Comput*. 2005;9: 653–668. doi:10.1109/TEVC.2005.856210
- 842 32. Hamann H. Evolution of Collective Behaviors by Minimizing Surprise. 14th Int Conf Synth Simul Living Syst
843 (ALIFE 2014). 2014; 344–351. doi:10.1145/2739482.2768497
- 844 33. Garnier S, Hamann H, Montes M, Christine DO, Eds TS, Hutchison D. Swarm Intelligence. In: Gerhard Goos,
845 Hartmanis J, Leeuwen J van, editors. LNCS 8667. Brussels; 2014.
- 846 34. Ishiwata H, Noman N, Iba H. Emergence of Cooperation in a Bio-inspired Multi-agent System. *Lecture Notes in*
847 *Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*
848 *Bioinformatics)*. 2010. pp. 364–374. doi:10.1007/978-3-642-17432-2_37
- 849 35. Reid CR, Lutz MJ, Powell S, Kao AB, Couzin ID, Garnier S. Army ants dynamically adjust living bridges in
850 response to a cost-benefit trade-off. *Proc Natl Acad Sci U S A*. 2015;112: 15113–8.
851 doi:10.1073/pnas.1512241112
- 852 36. Joshi NJ, Tononi G, Koch C. The Minimal Complexity of Adapting Agents Increases with Fitness. *PLoS Comput*
853 *Biol*. 2013;9. doi:10.1371/journal.pcbi.1003111
- 854 37. Sheneman L, Hintze A. Evolving autonomous learning in cognitive networks. *Sci Rep*. Springer US; 2017; 1–11.
855 doi:10.1038/s41598-017-16548-2
- 856 38. Beer RD, Williams PL. Information Processing and Dynamics in Minimally Cognitive Agents. *Cogn Sci*.
857 2015;39: 1–38. doi:10.1111/cogs.12142
- 858 39. Lizier JT, Prokopenko M, Zomaya AY. A Framework for the Local Information Dynamics of Distributed
859 Computation in Complex Systems. 2014. pp. 115–158. doi:10.1007/978-3-642-53734-9_5
- 860 40. Zenil H. Compression-based investigation of the dynamical properties of cellular automata and other systems.
861 *Arxiv Prepr arXiv09104042*. 2009; 1–25. Available: <http://arxiv.org/abs/0910.4042>
- 862 41. Clifford Bohm, Nitash C. G. AH. MABE (Modular Agent Based Evolver): A framework for digital evolution

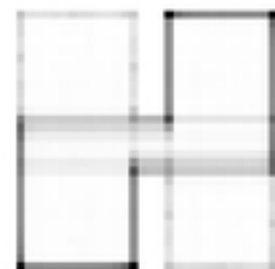
- 863 research. Proceedings of the European Conference on Artificial Life. MIT Press; 2017. pp. 76–83.
- 864 42. Hintze A, Edlund JA, Olson RS, Knoester DB, Schossau J, Albantakis L, et al. Markov Brains: A Technical
865 Introduction. 2017; Available: <http://arxiv.org/abs/1709.05601>
- 866 43. Tononi G. Integrated information theory. Scholarpedia. 2015;10: 4164. doi:10.4249/scholarpedia.4164
- 867 44. Marshall W, Kim H, Walker SI, Tononi G, Albantakis L. How causal analysis can reveal autonomy in models of
868 biological systems. Philos Trans R Soc A Math Phys Eng Sci. 2015; 1–22. doi:10.1098/not
- 869 45. Albantakis L. A Tale of Two Animats: What does it take to have goals? 2017;
- 870



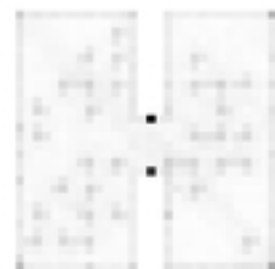
2-dimensional environment



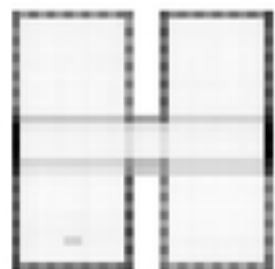
$\langle G_{0.50}^{10k} \rangle$
 $\langle EF \rangle = 3.13$



$\langle G_{3sides}^{10k} \rangle$
 $\langle EF \rangle = 3.61$



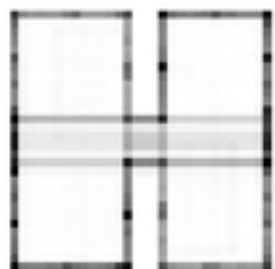
$\langle G_{!agent}^{10k} \rangle$
 $\langle EF \rangle = 0.18$



$\langle G_{!feedback}^{10k} \rangle$
 $\langle EF \rangle = 1.97$

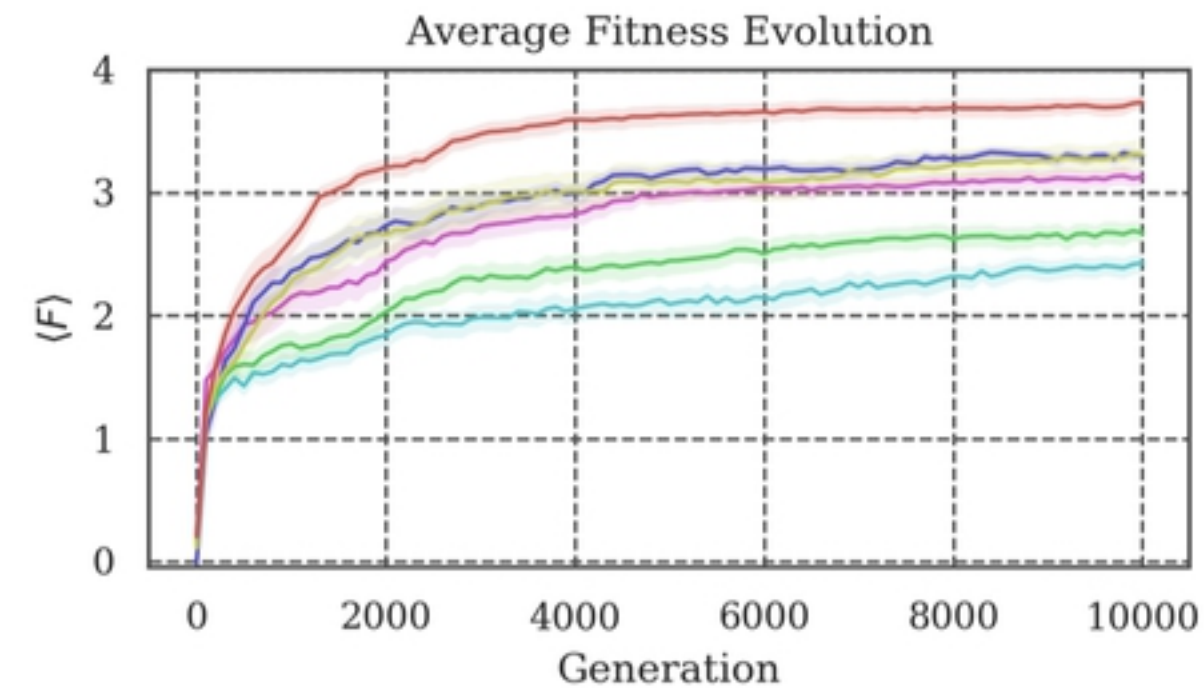


$\langle G_{w=a}^{10k} \rangle$
 $\langle EF \rangle = 0.18$



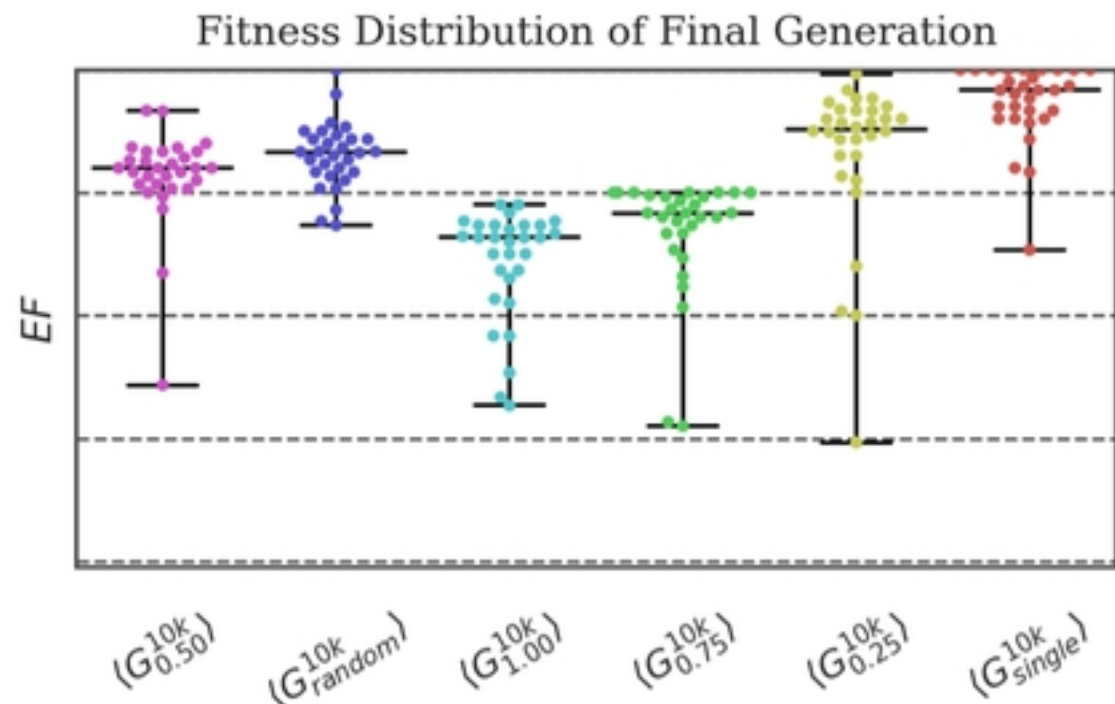
$\langle G_{random}^{10k} \rangle$
 $\langle EF \rangle = 3.31$

Fig 1



— $\langle G_{0.50} \rangle$
 — $\langle G_{random} \rangle$
 — $\langle G_{1.00} \rangle$
 — $\langle G_{0.75} \rangle$
 — $\langle G_{0.25} \rangle$
 — $\langle G_{single} \rangle$

(a)



(b)

Fig 2

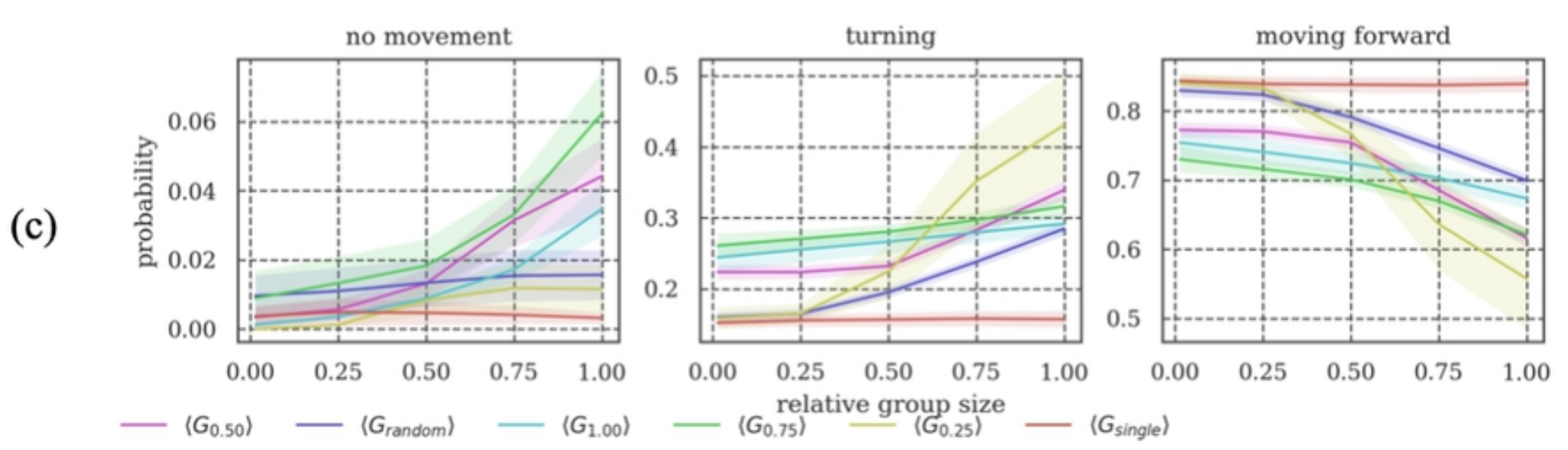
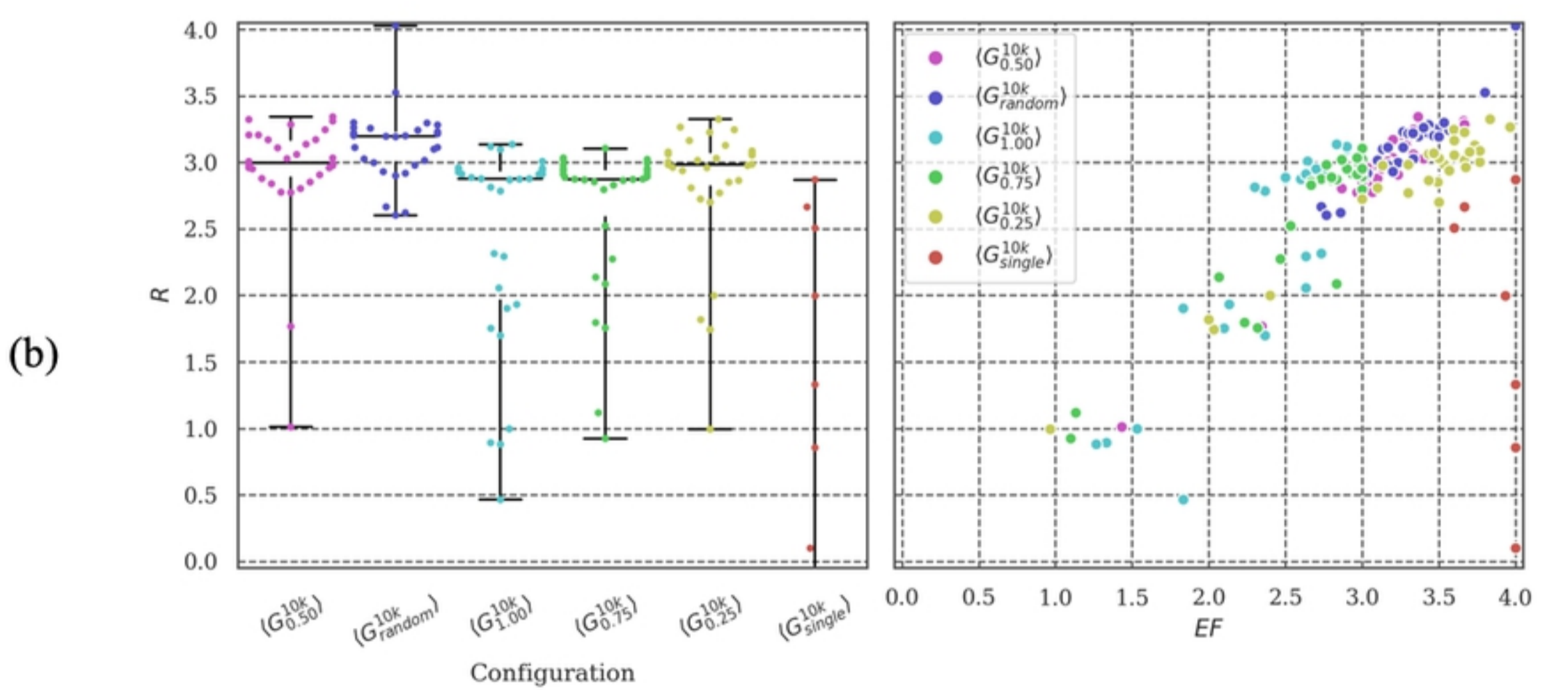
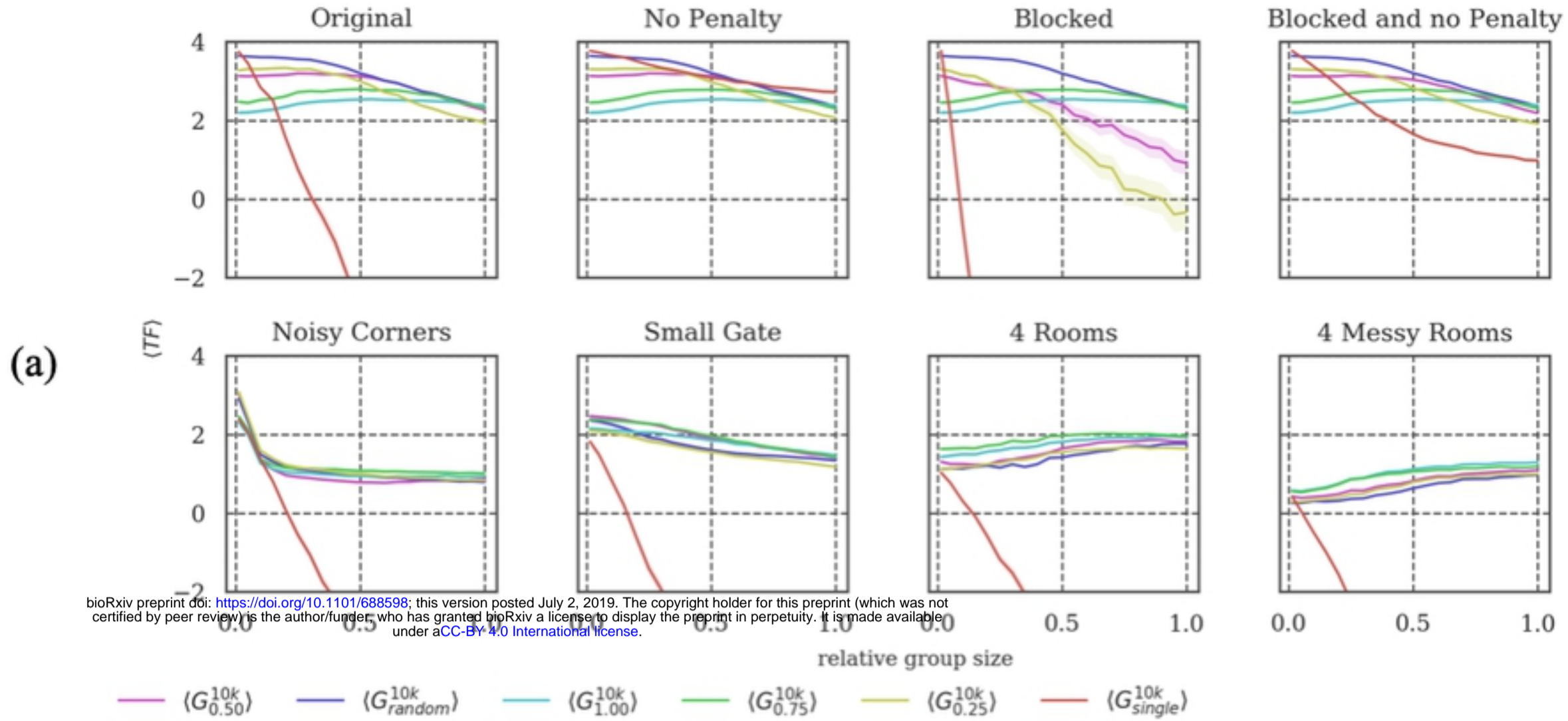
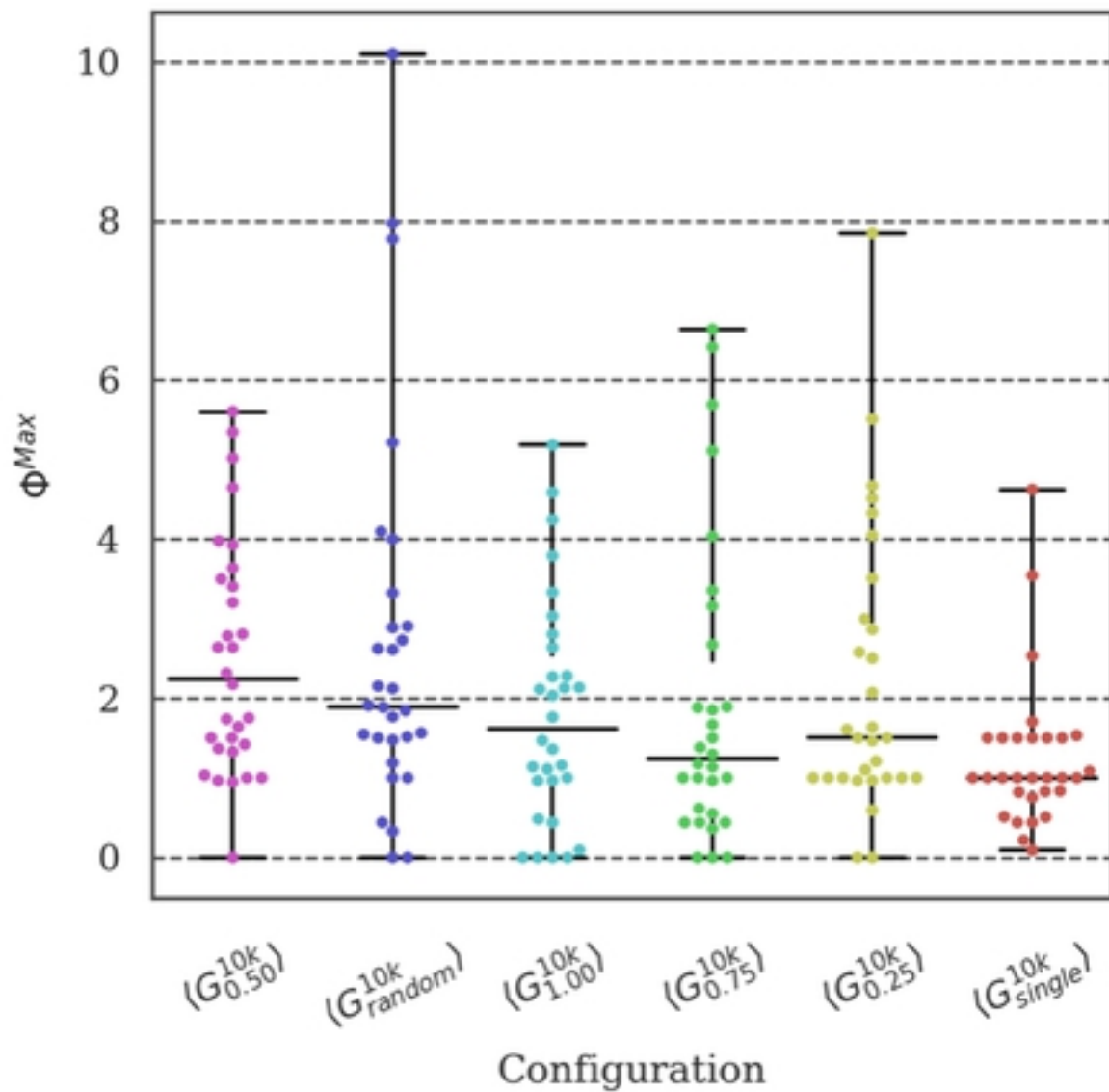
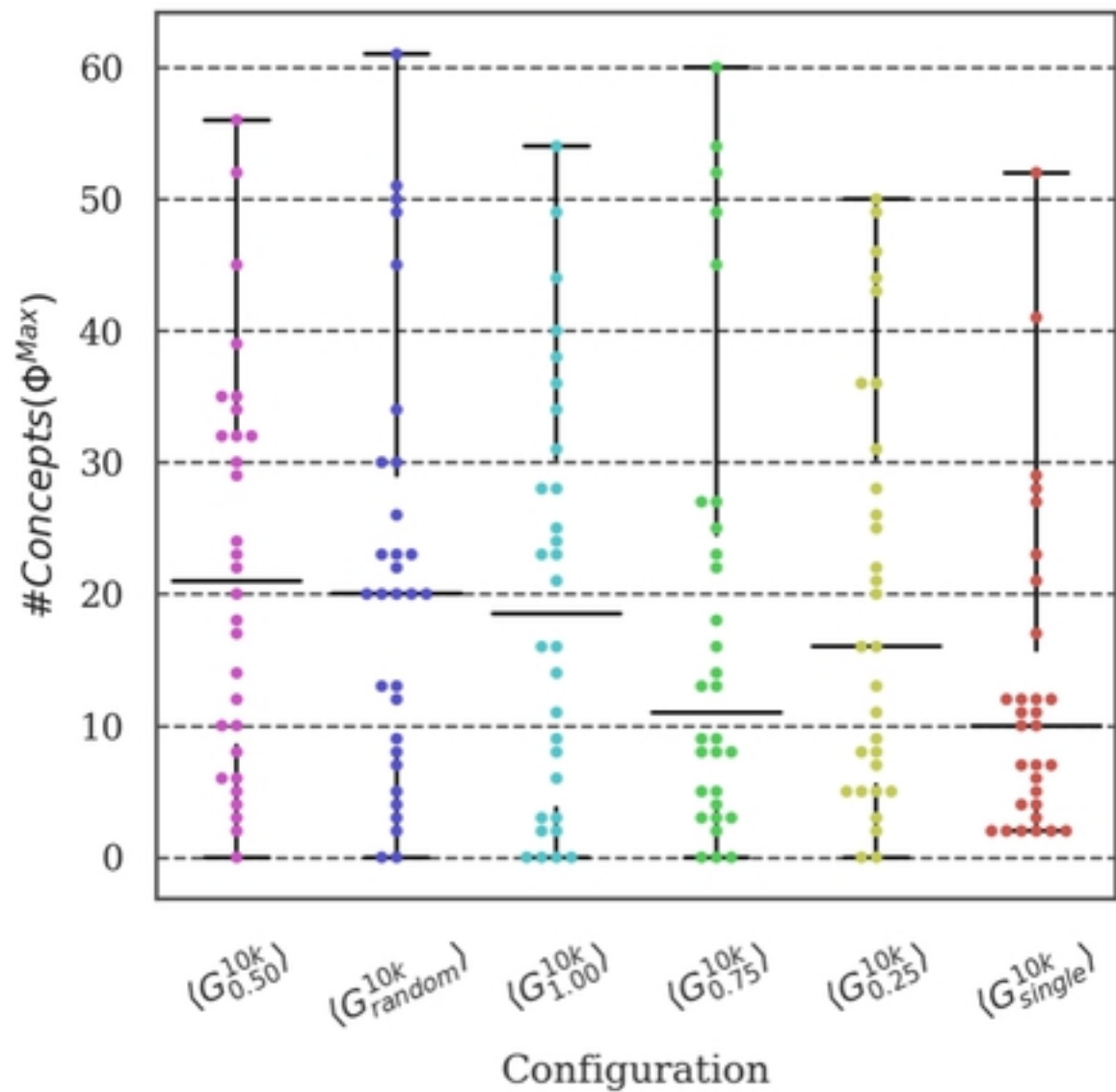


Fig 3



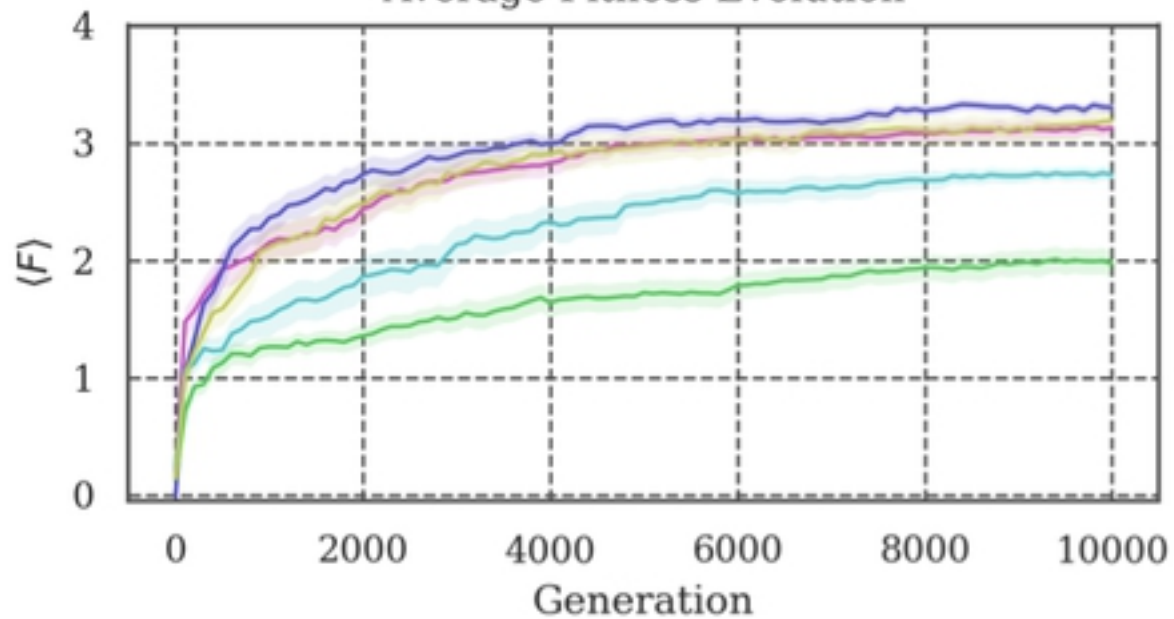
(a)



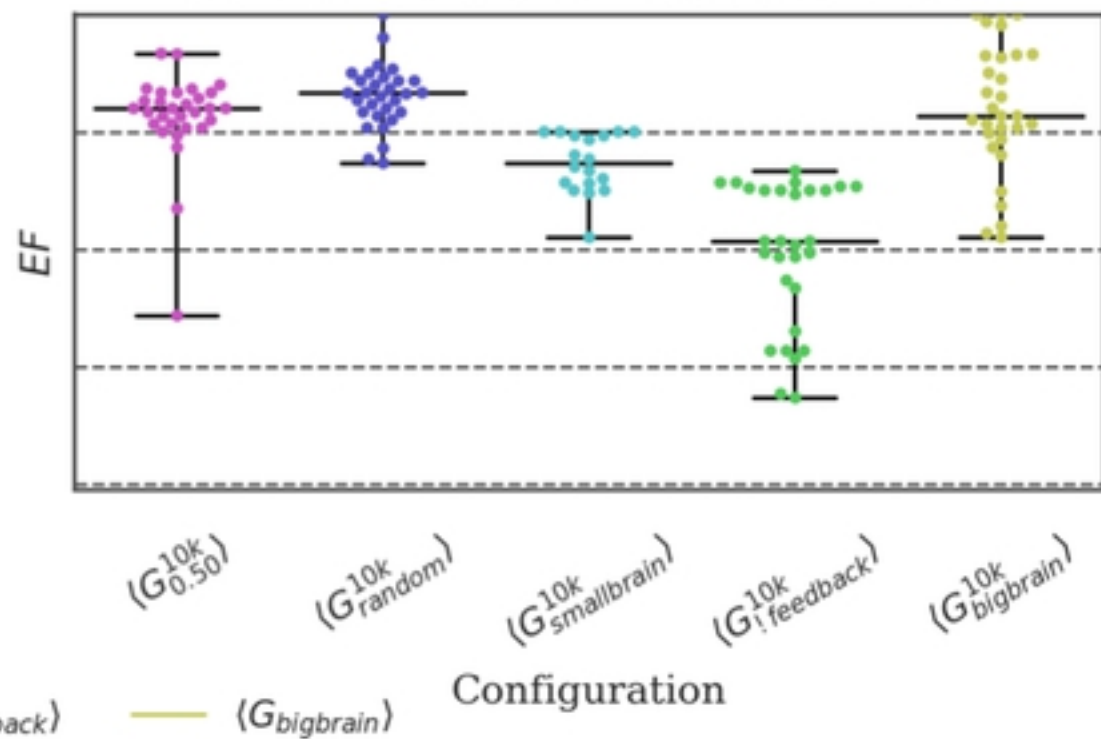
(b)

Fig 4

Average Fitness Evolution



Fitness Distribution of Final Generation



(a)

(b)

Fig 5

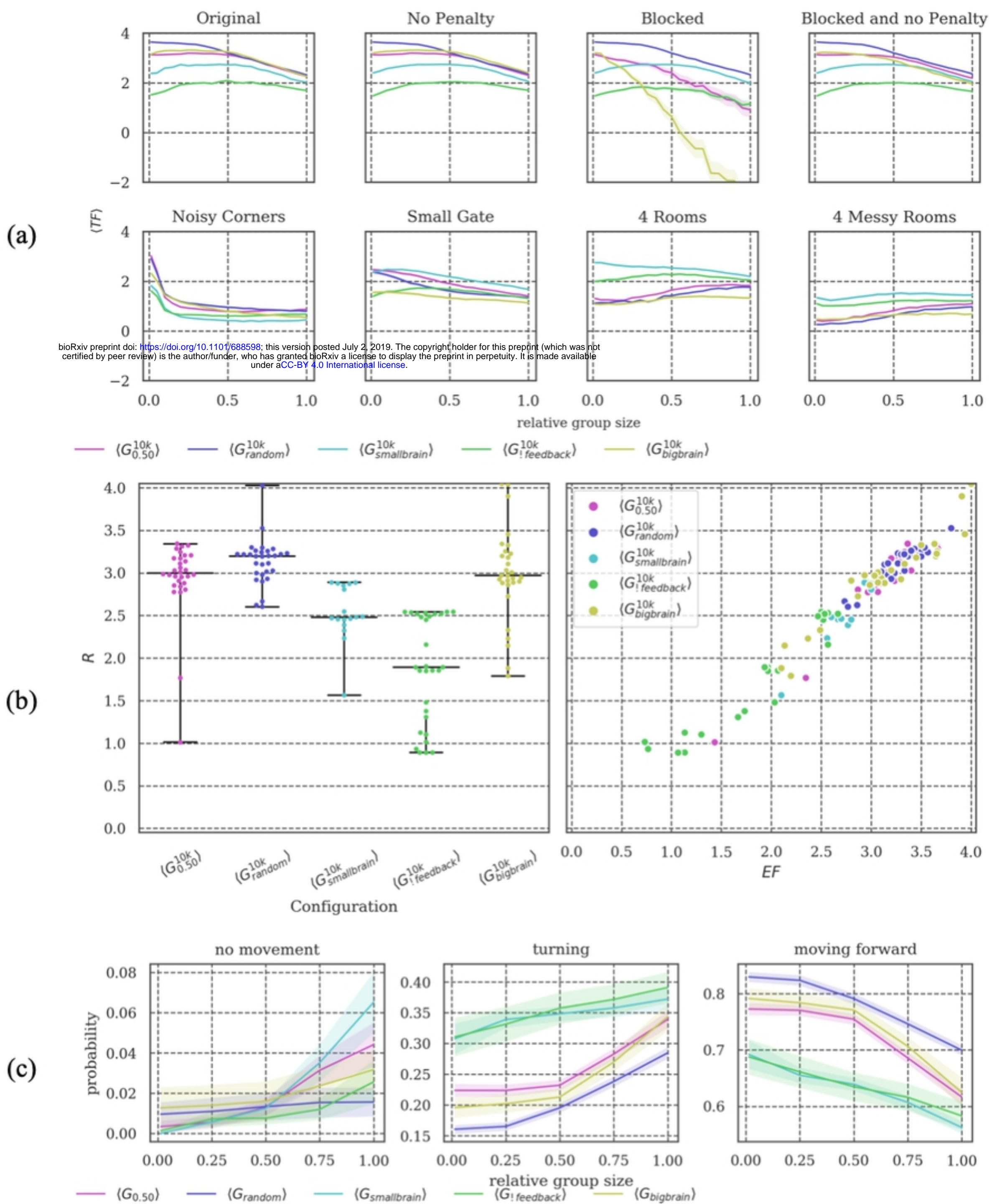
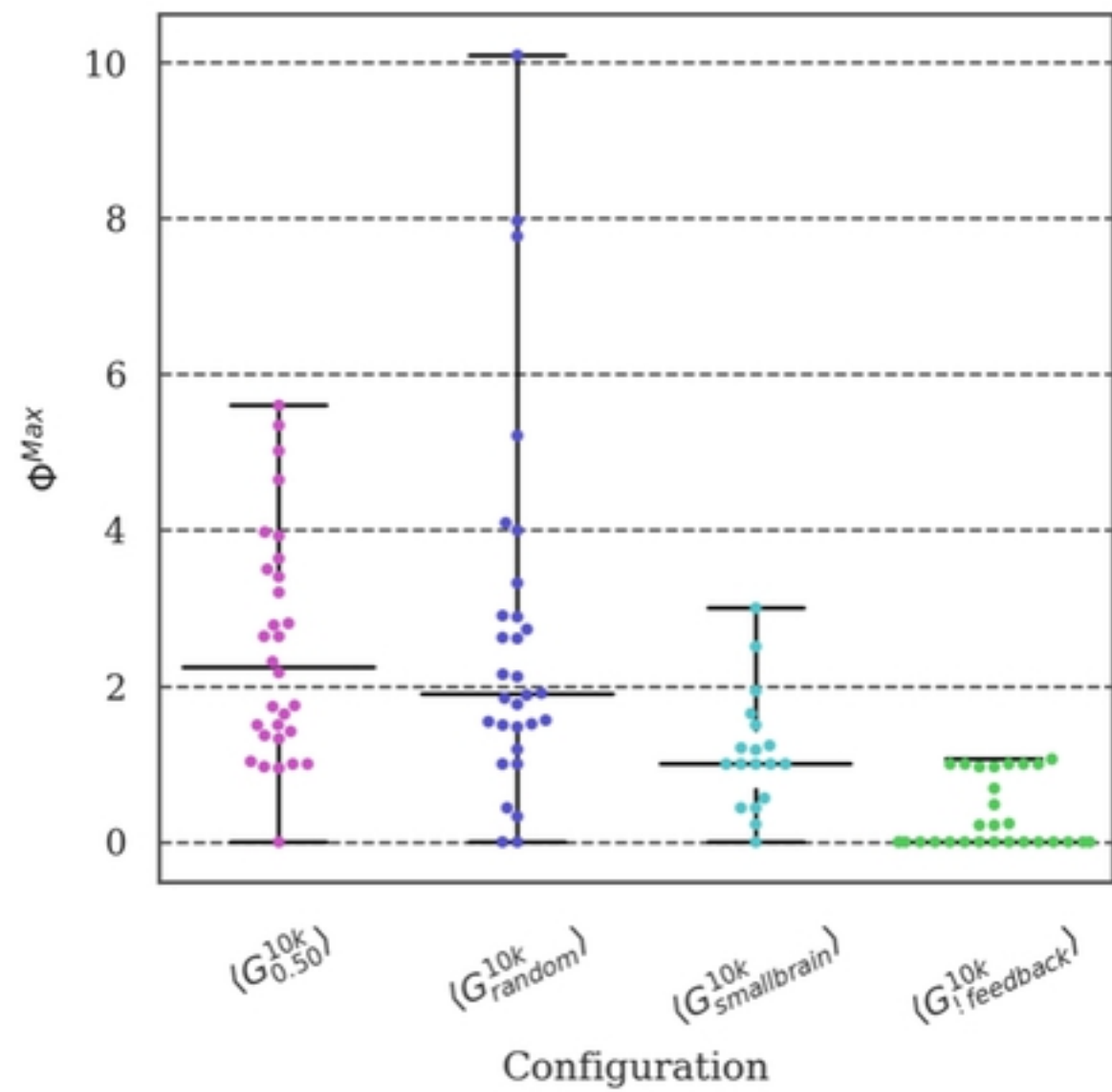
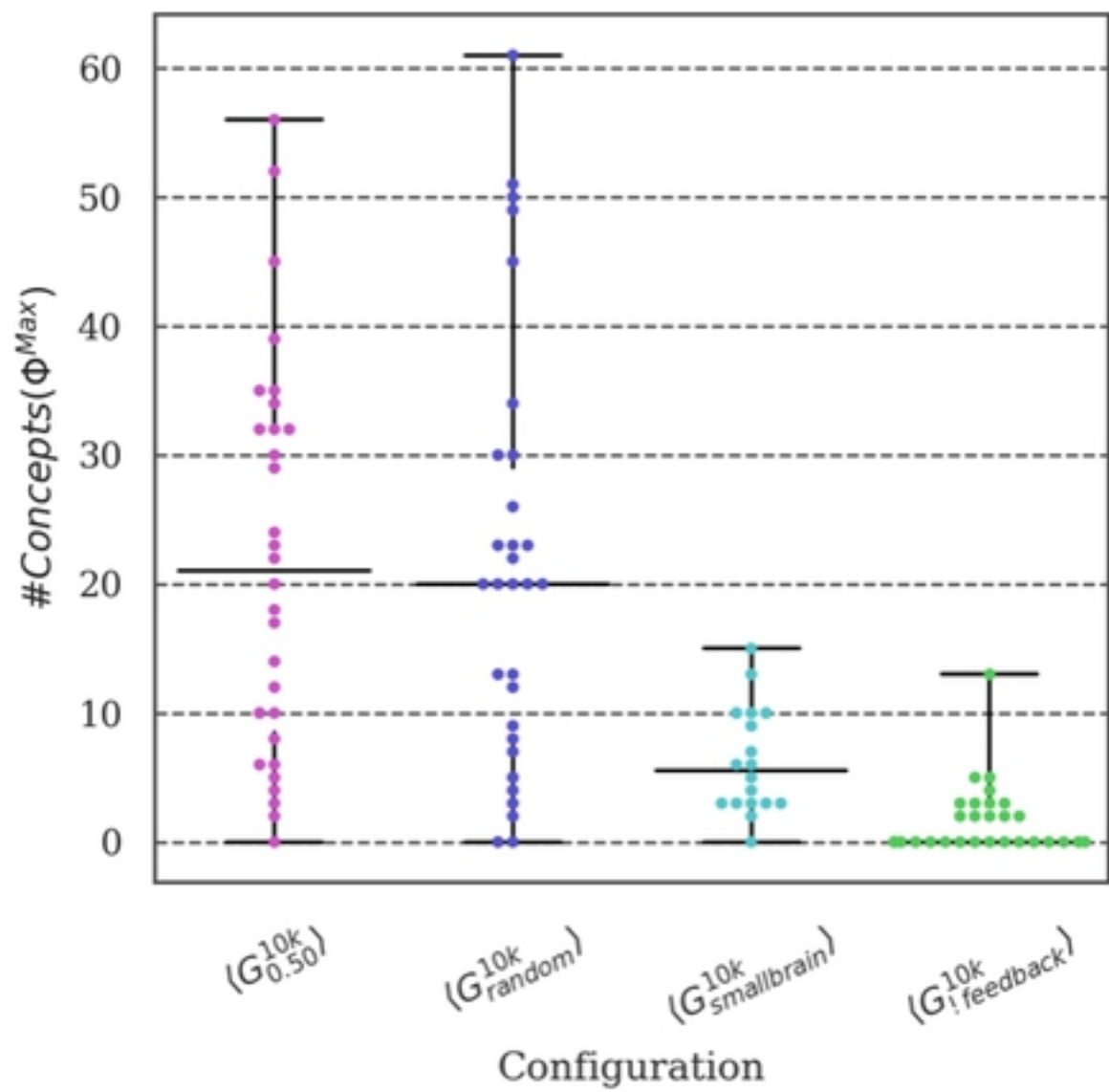


Fig 6



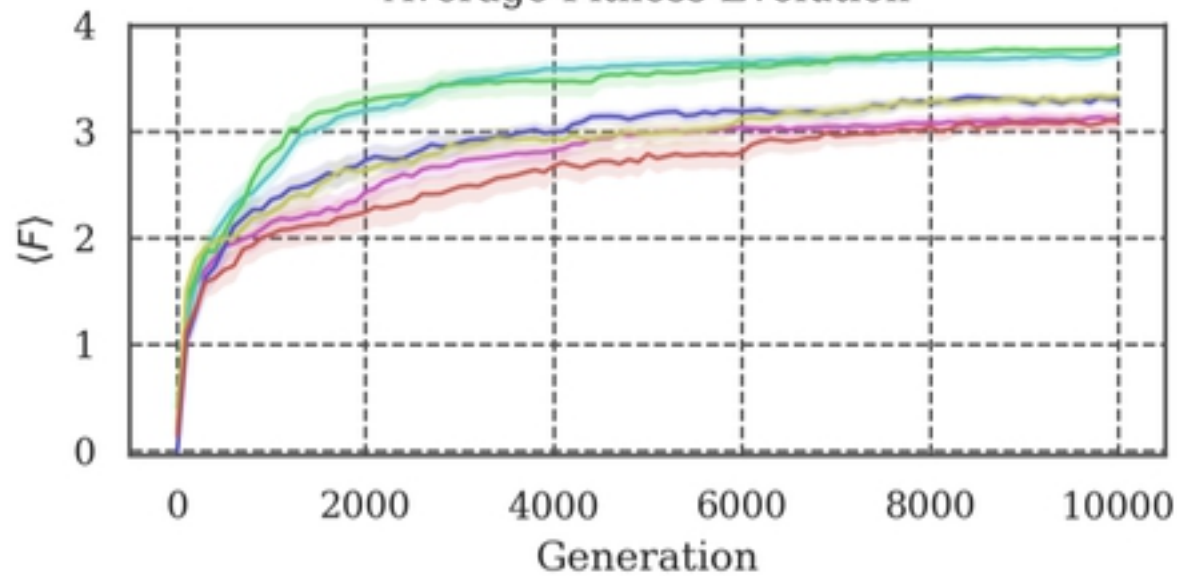
(a)



(b)

Fig 7

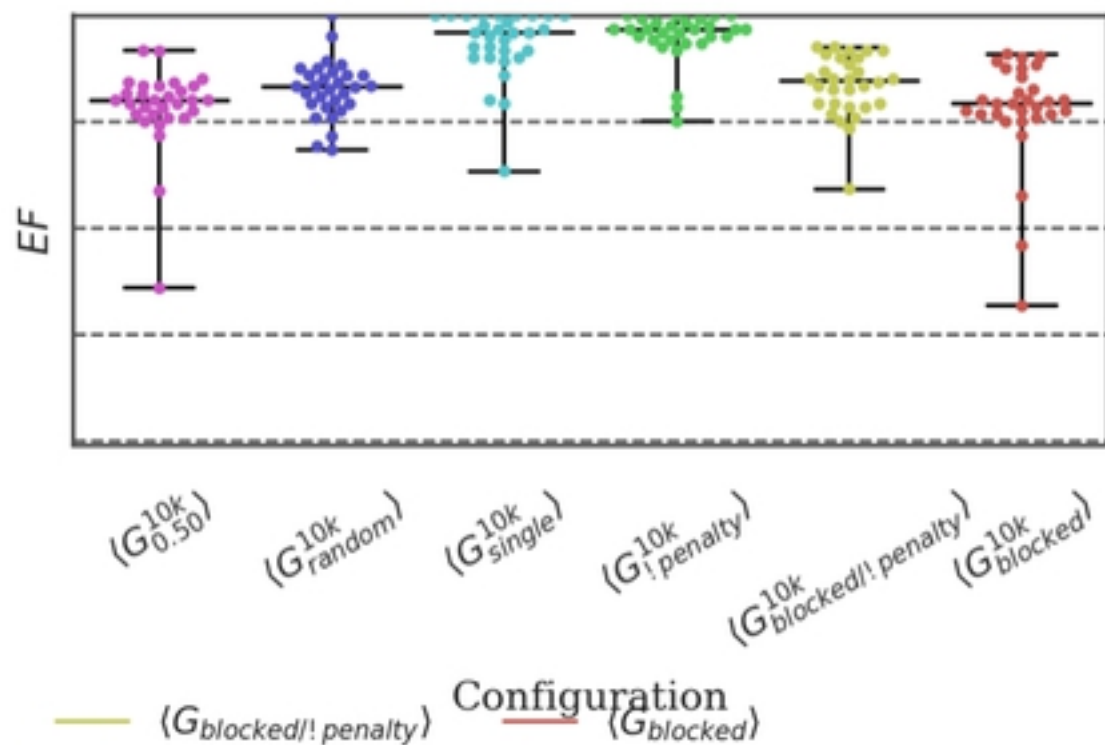
Average Fitness Evolution



— $\langle G_{0.50} \rangle$ — $\langle G_{\text{random}} \rangle$ — $\langle G_{\text{single}} \rangle$ — $\langle G_{! \text{penalty}} \rangle$

(a)

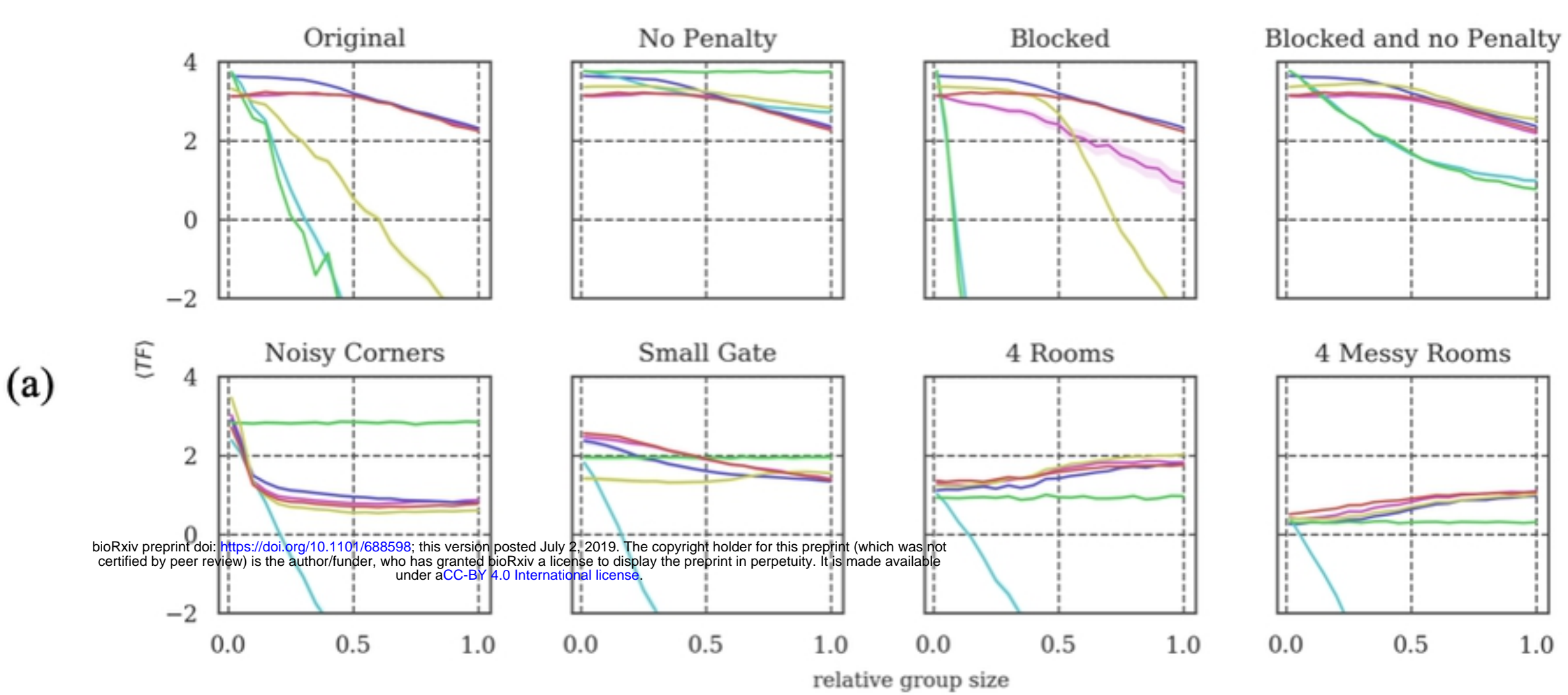
Fitness Distribution of Final Generation



— $\langle G_{\text{blocked}!/ \text{penalty}} \rangle$ — $\langle G_{\text{blocked}} \rangle$

(b)

Fig 8



$\langle G_{0.50}^{10k} \rangle$
 $\langle G_{random}^{10k} \rangle$
 $\langle G_{single}^{10k} \rangle$
 $\langle G_{!penalty}^{10k} \rangle$
 $\langle G_{blocked!penalty}^{10k} \rangle$
 $\langle G_{blocked}^{10k} \rangle$

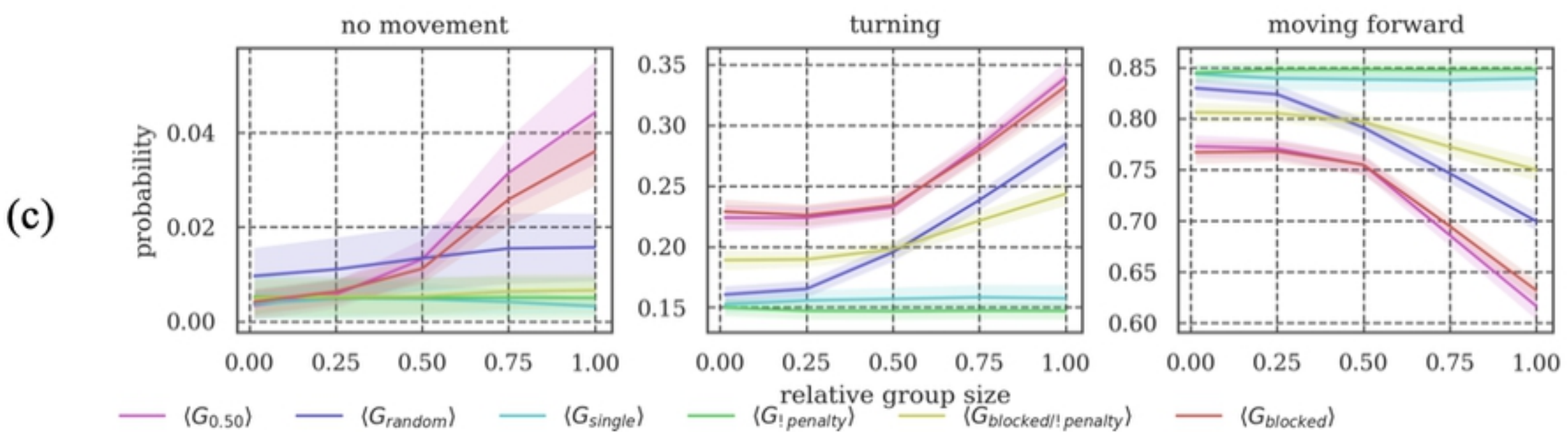
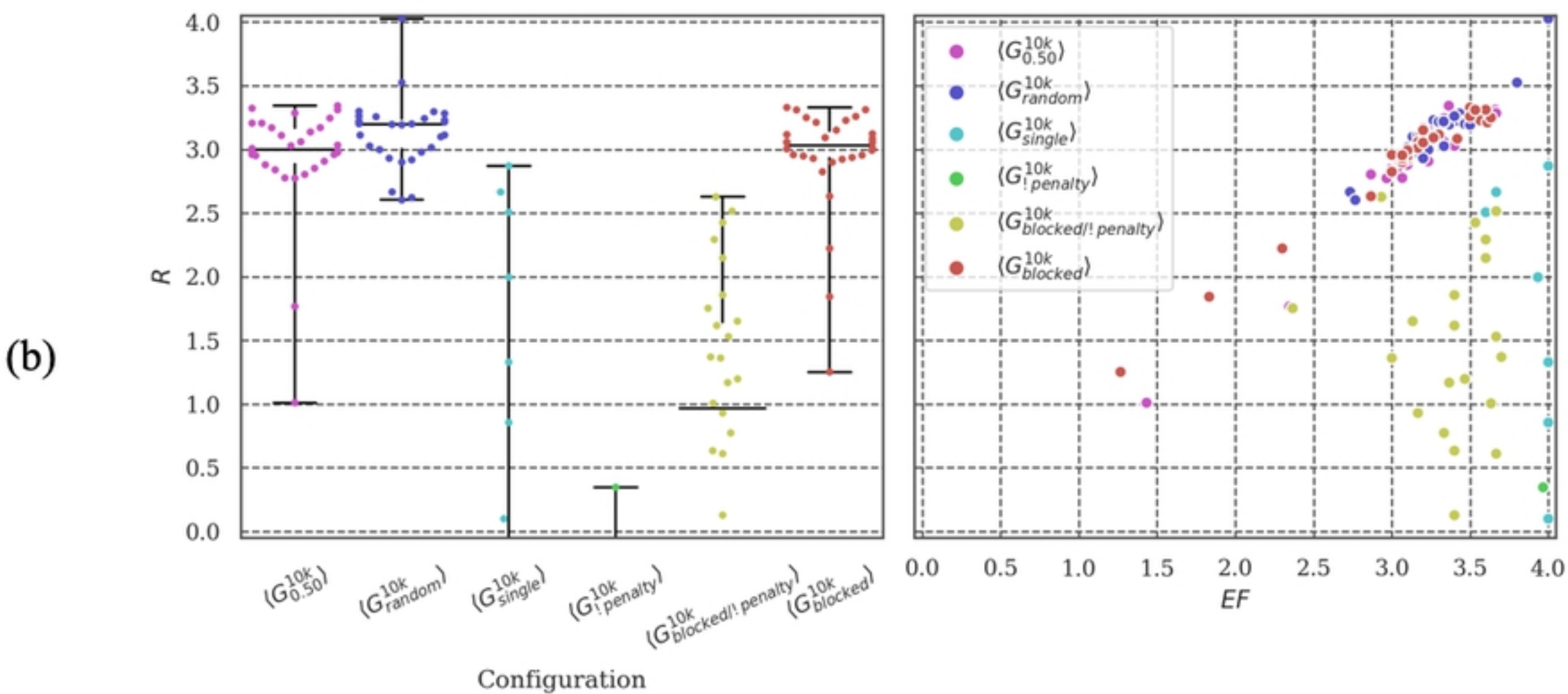
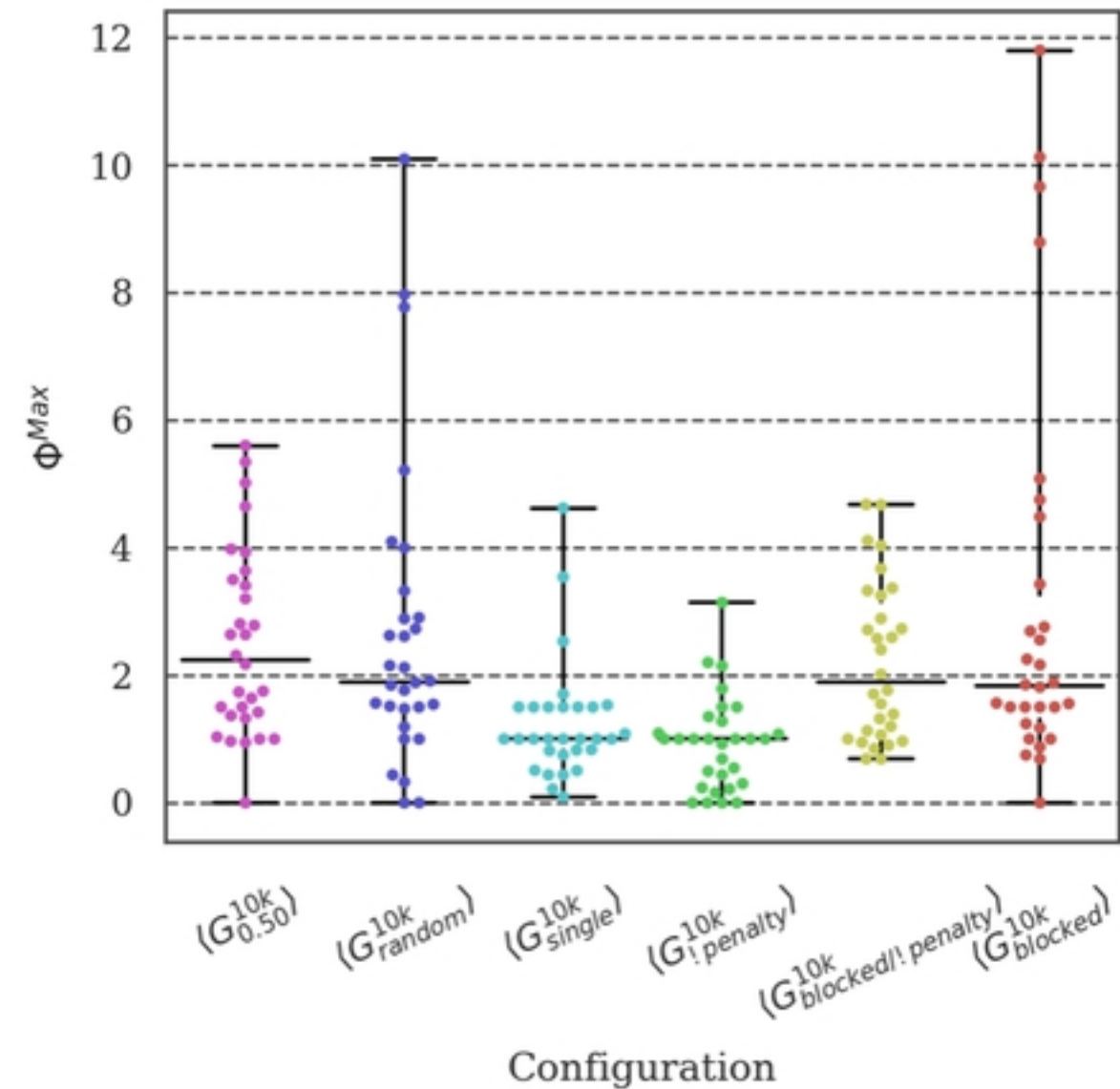
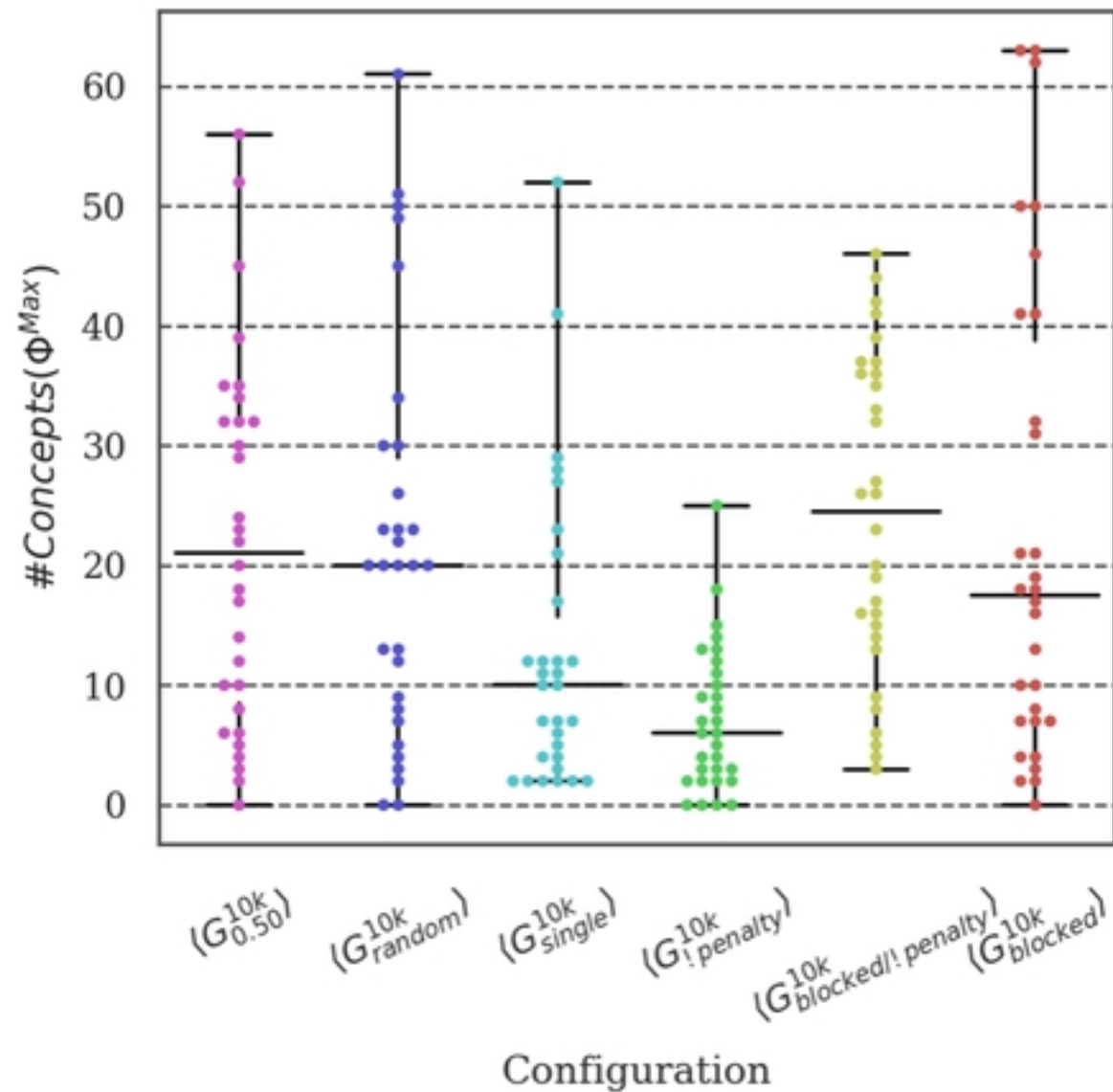


Fig 9

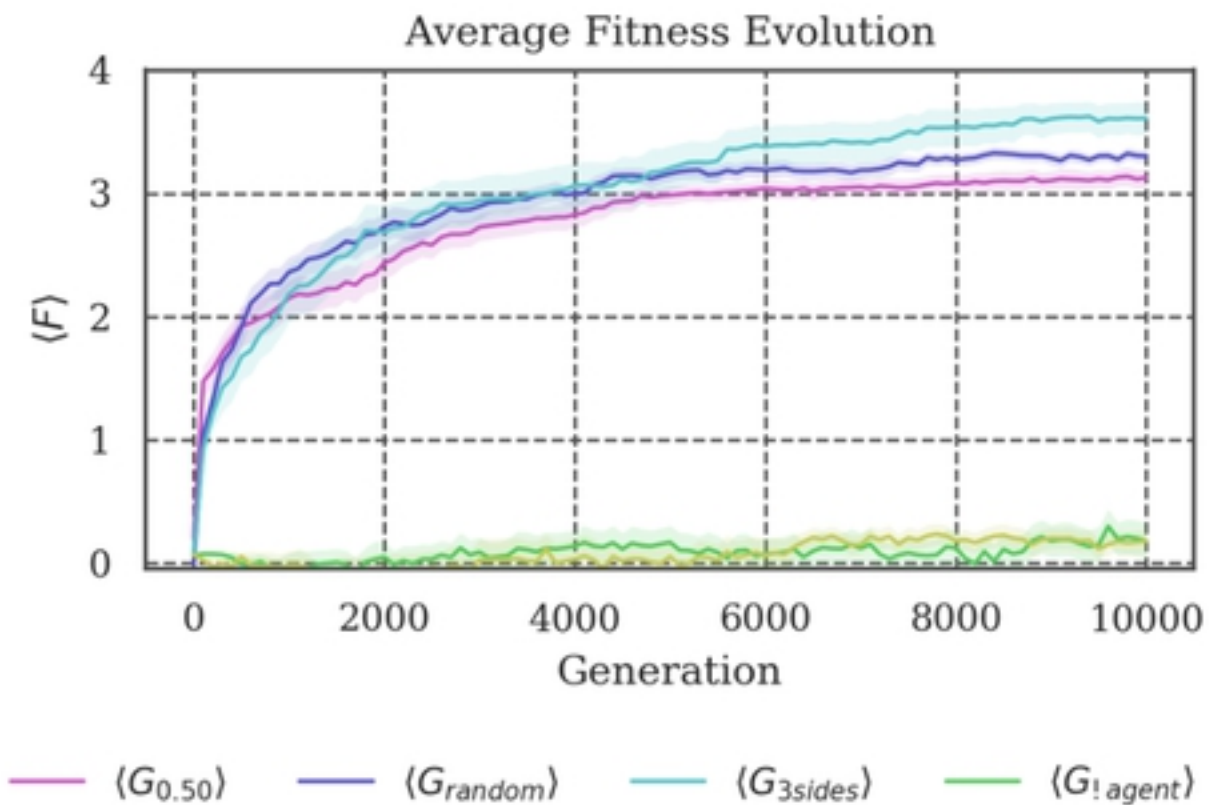


(a)

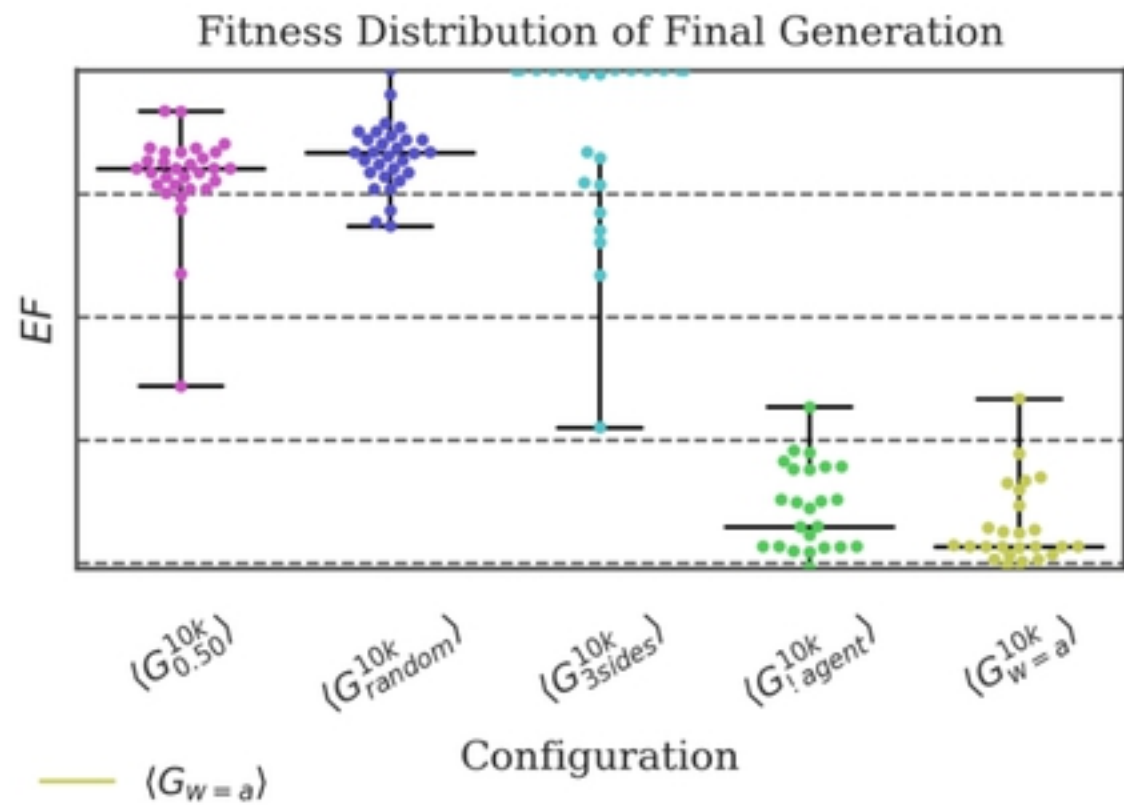


(b)

Fig 10

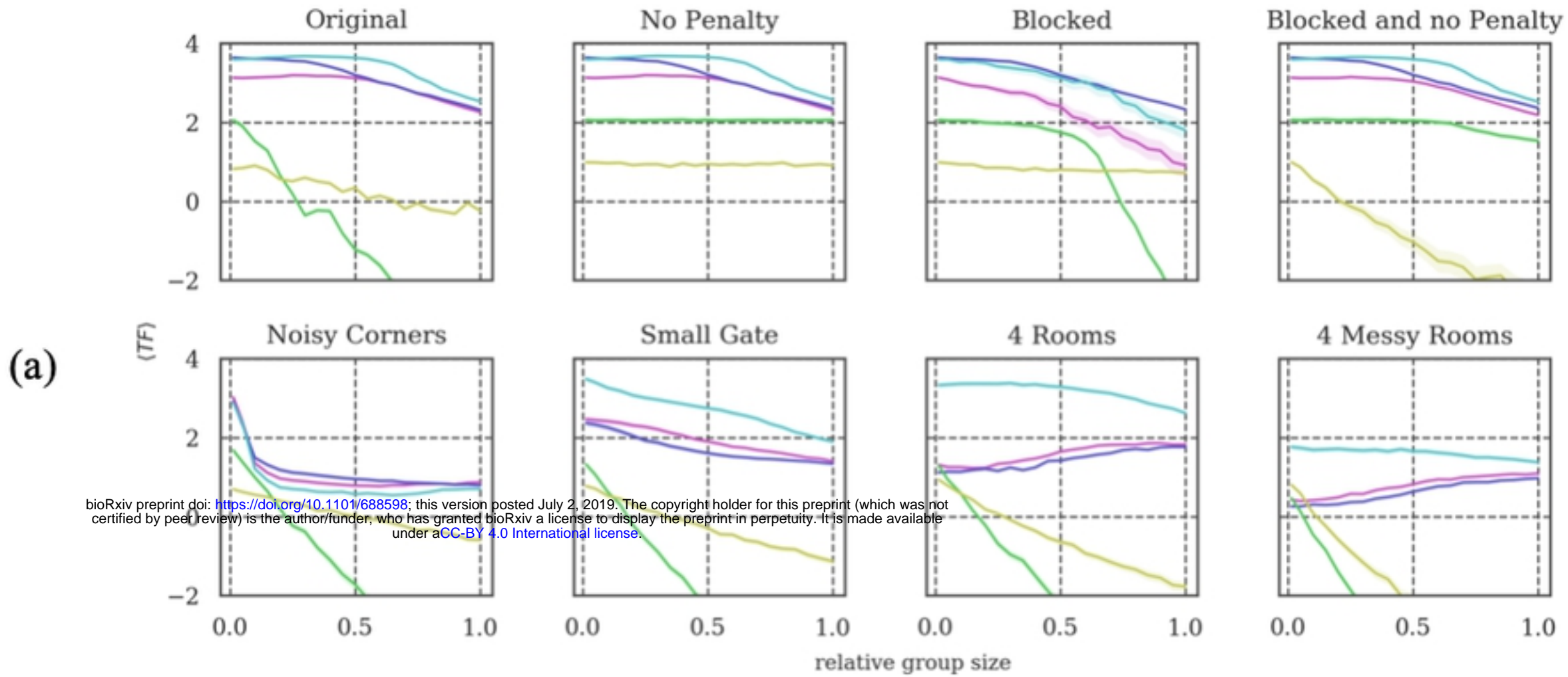


(a)



(b)

Fig 11



bioRxiv preprint doi: <https://doi.org/10.1101/688598>; this version posted July 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

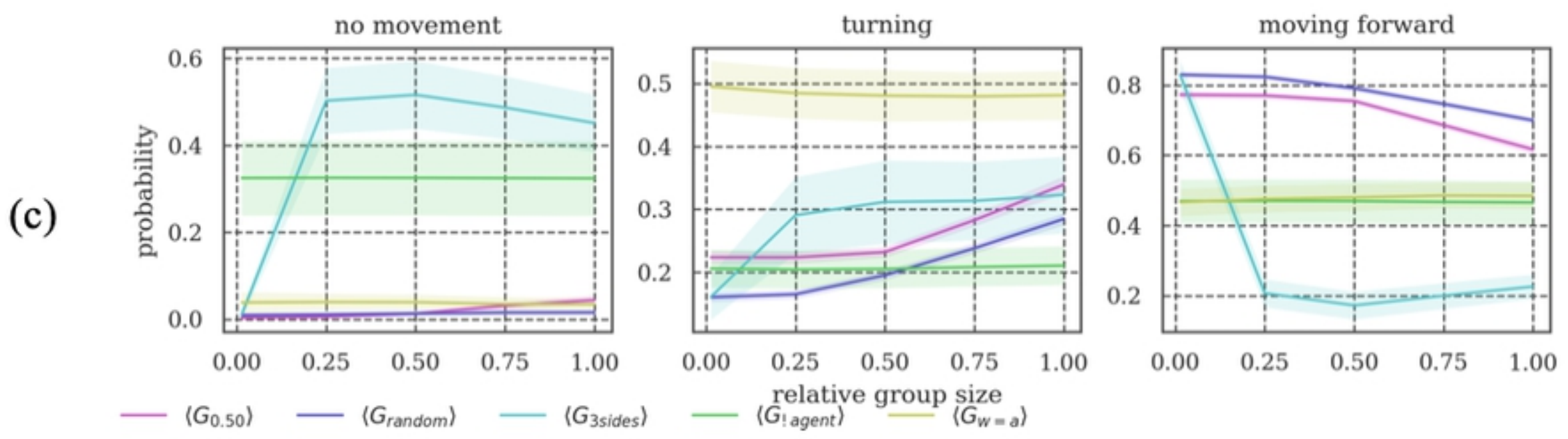
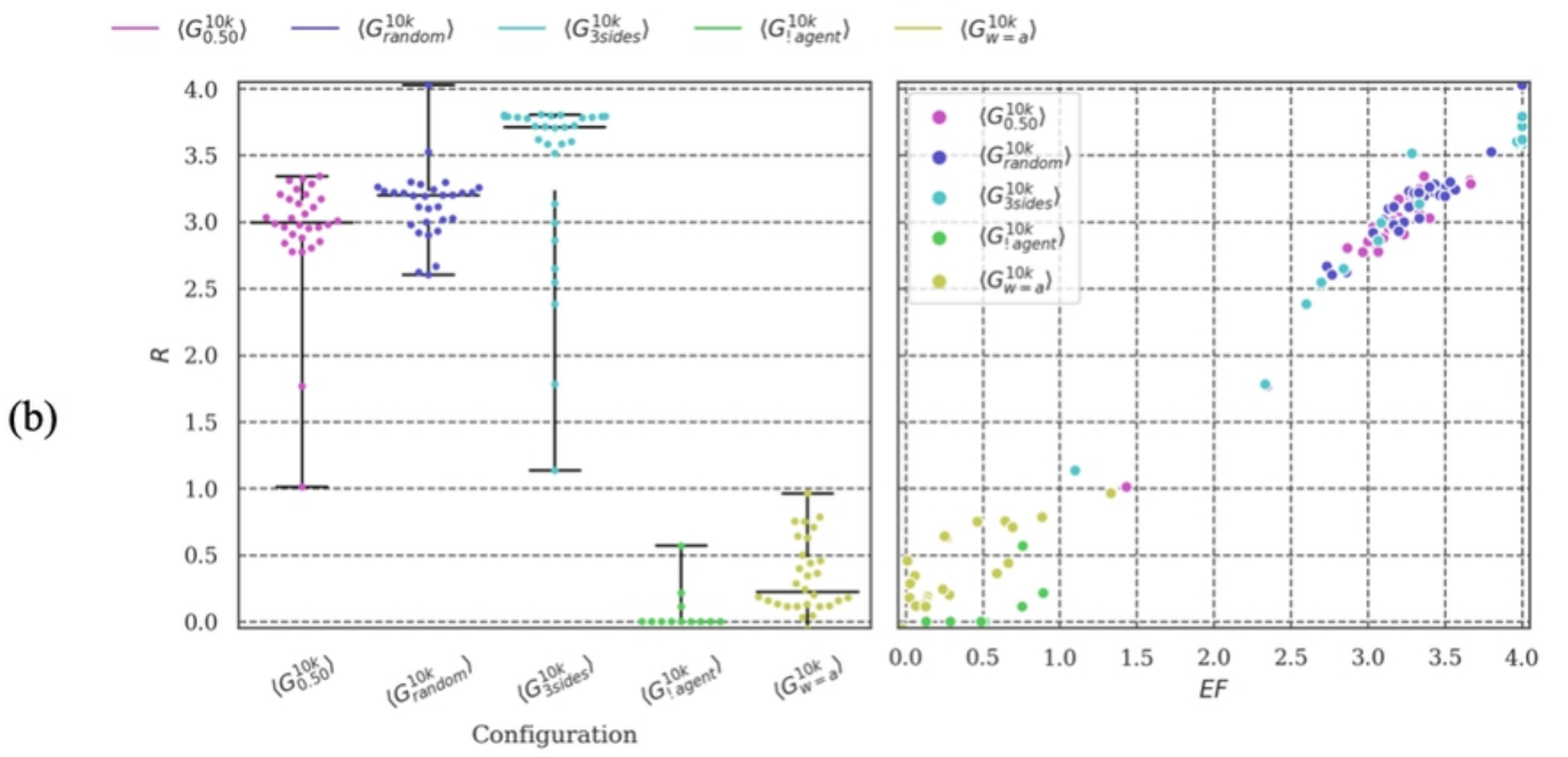
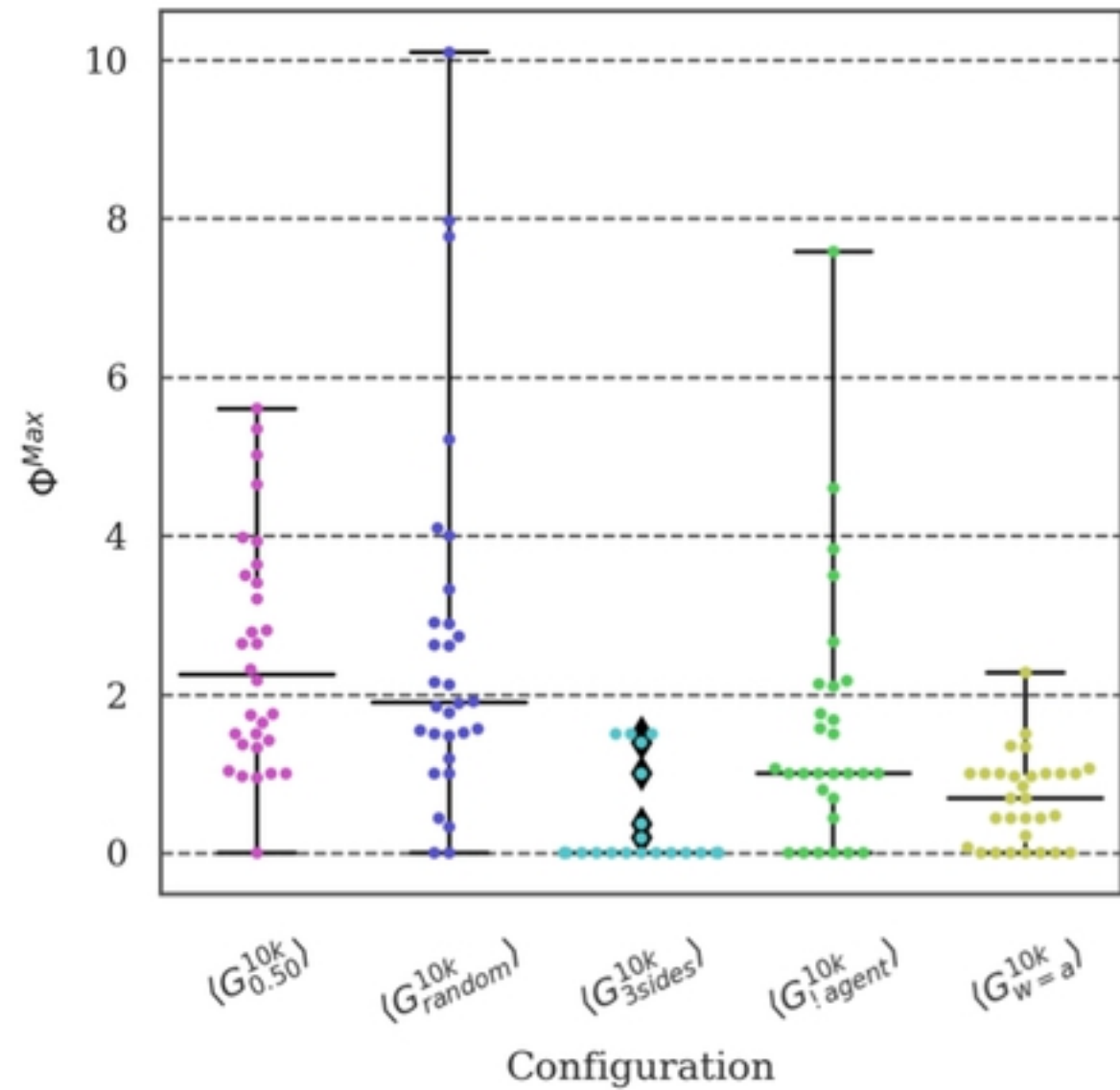
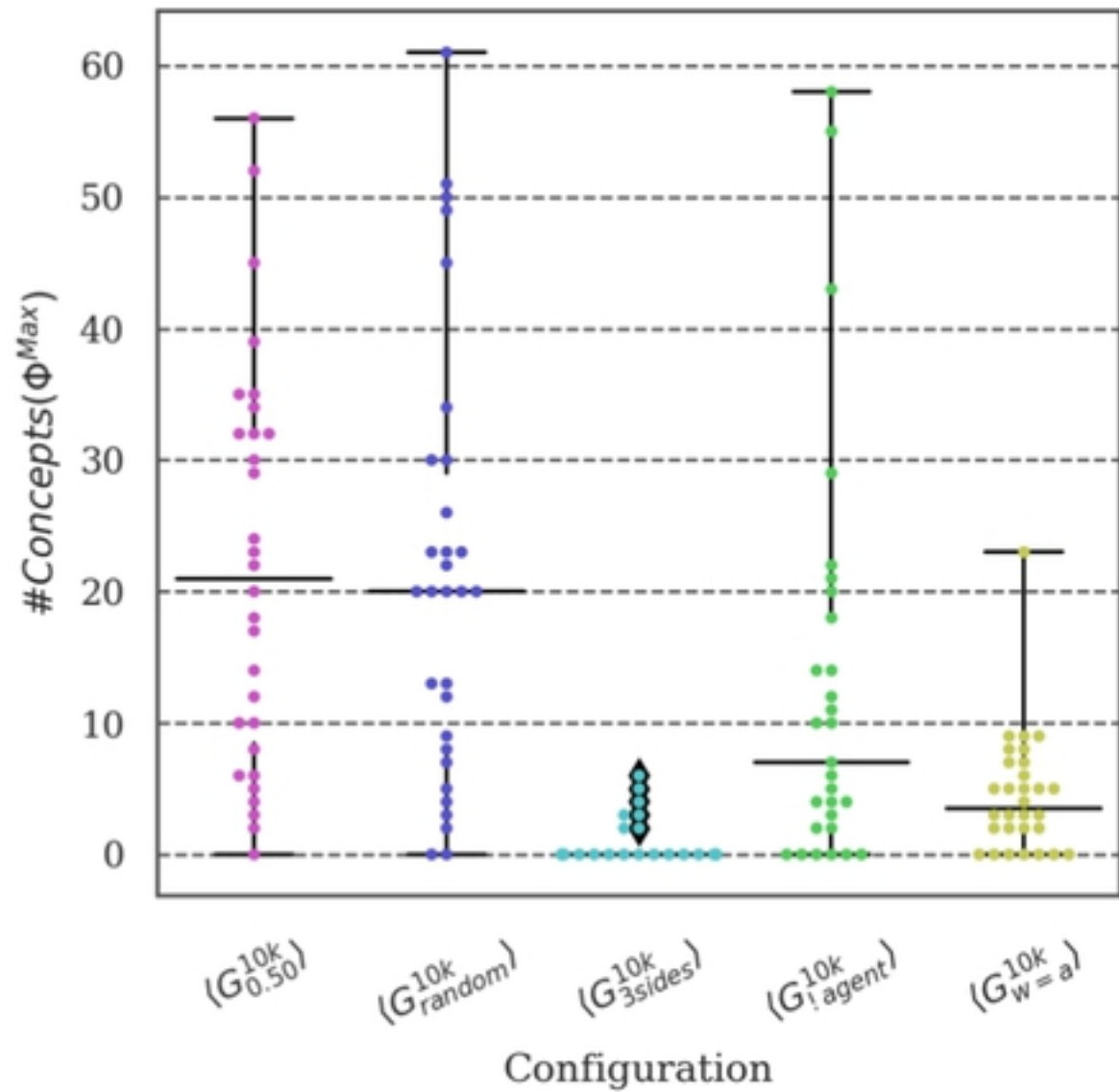


Fig 12



(a)



(b)

Fig 13

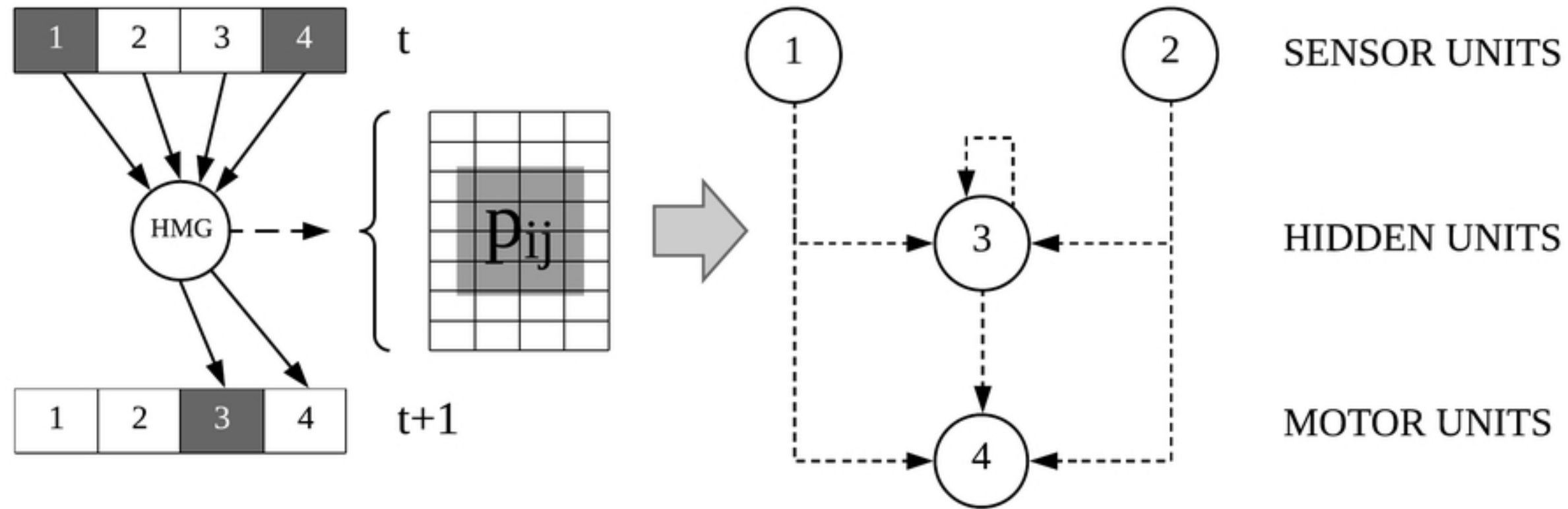


Fig 14

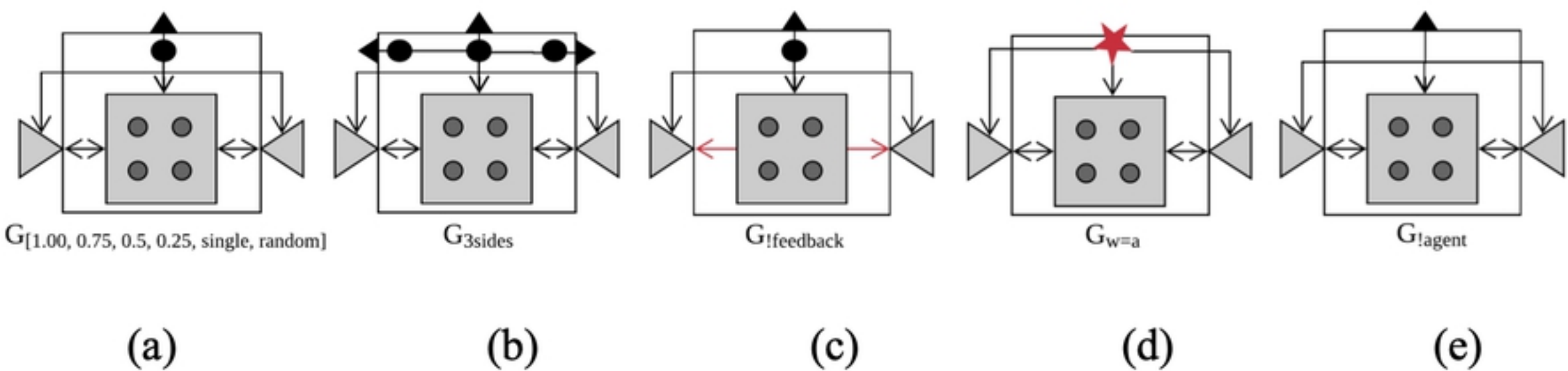
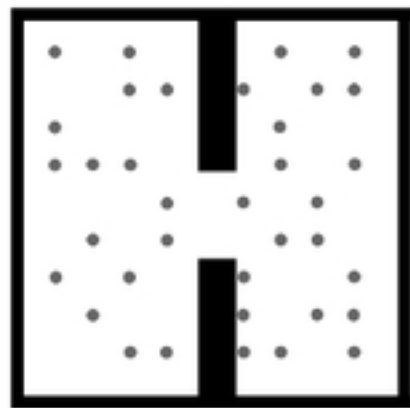
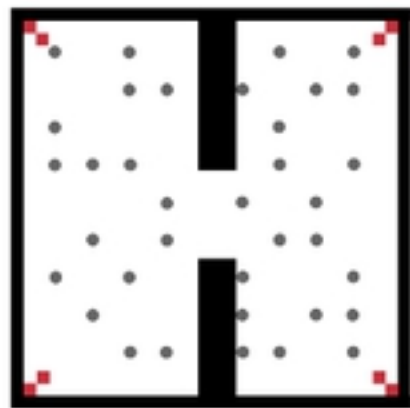


Fig 15



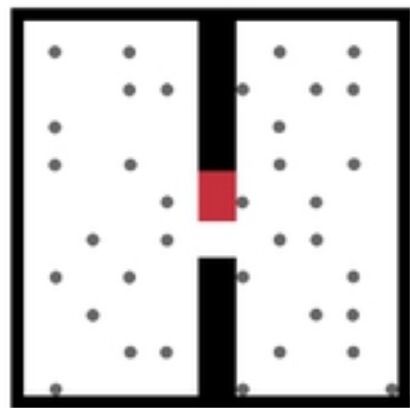
Original

(a)



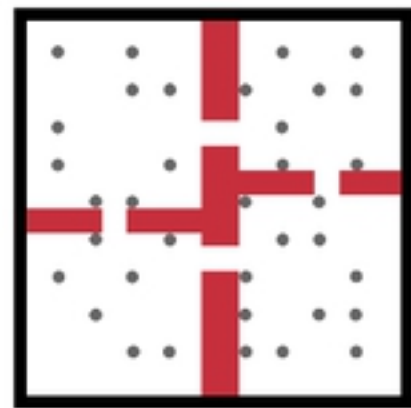
Noisy Corners

(b)



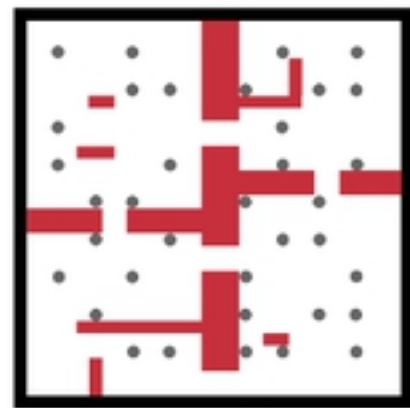
Small Gate

(c)



4 Rooms

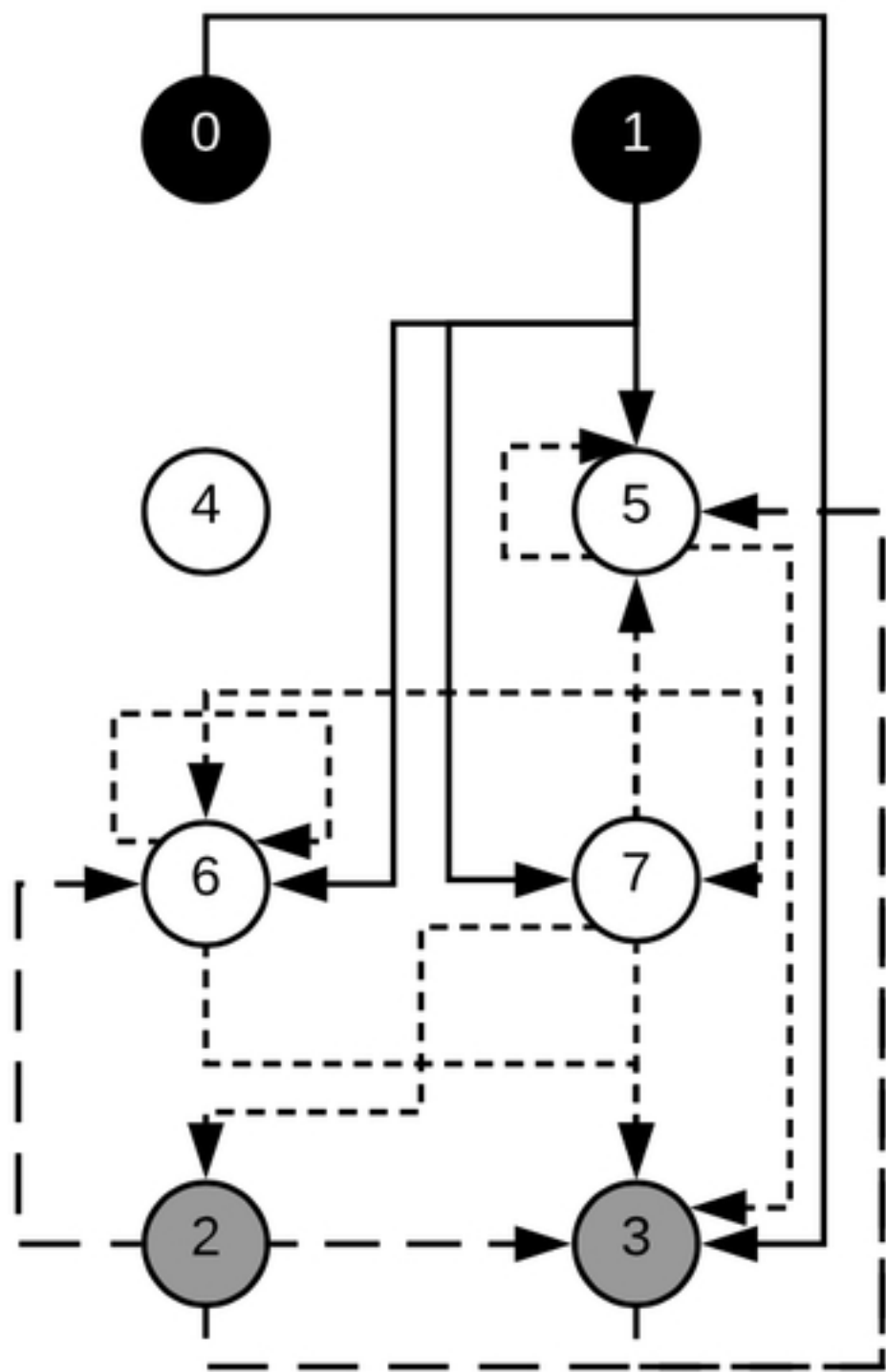
(d)



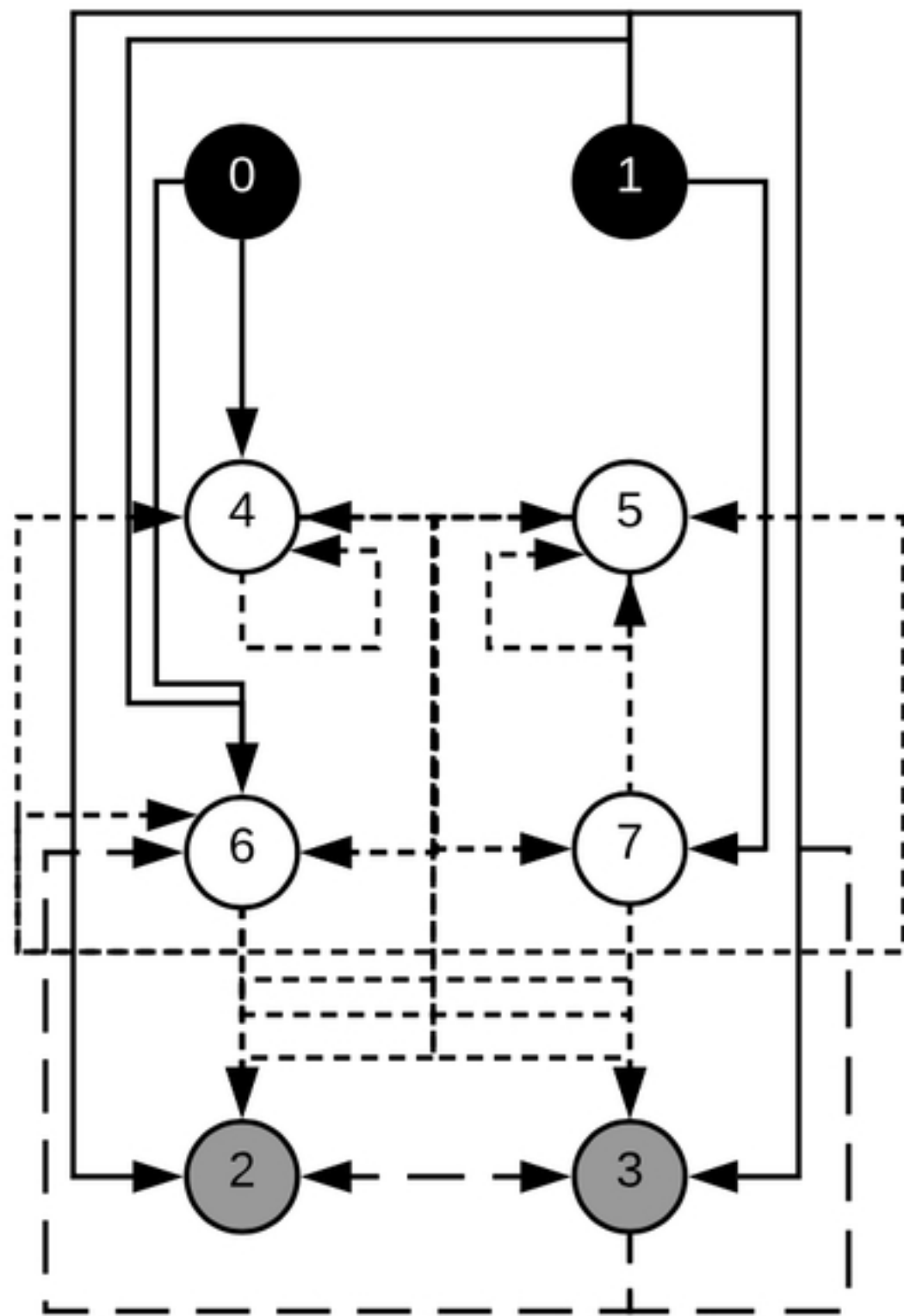
4 Messy Rooms

(e)

Fig 16



(a)



(b)

Fig S1