

TITLE: Structure and influence in an interconnected world: neurocomputational mechanism of real-time distributed learning on social networks

AUTHORS: Yaomin Jiang^a, Qingtian Mi^a, Lusha Zhu^{a*}

^a Peking University, Beijing, 100871, China

*CORRESPONDING AUTHORS:

Lusha Zhu

52 Haidian Road,

Wang Kezhen Bldg, Peking University

Beijing, 100871, China

Email: lushazhu@pku.edu.cn

1 **Abstract**

2 Many social species are embedded on social networks, including our own. The structure of
3 social networks shapes our decisions by constraining what information we learn and from
4 whom. But how does the brain incorporate social network structures into learning and decision-
5 making processes, and how does learning in networked environments differ from learning from
6 isolated partners? Combining a real-time distributed learning task with computational
7 modeling, fMRI, and social network analysis, we investigated the process by which humans
8 learn from observing others' decisions on 7-node networks with varying topological structures.
9 We show that learning on social networks can be realized by means similar to the well-
10 established reinforcement learning algorithm, supported by an action prediction error encoded
11 in the lateral prefrontal cortex. Importantly, learning is flexibly weighted toward well-
12 connected neighbors, according to activity in the dorsal anterior cingulate cortex, but only
13 insofar as neighbors' actions vary in their informativeness. These data suggest a
14 neurocomputational mechanism of network-dependent filtering on the sources of information,
15 which may give rise to biased learning and the spread of misinformation in an interconnected
16 society.

17 Social networks channel communication and route information transmission in human society^{1,2}. By
18 constraining what information we receive and from whom, the structure of social networks has
19 substantial impacts on how we form beliefs and make decisions, and how collective opinion and
20 behavior are shaped and propagated³⁻⁵. Although many studies have demonstrated the influence of
21 social networks on political⁶, economic⁷, and social activities⁸ at the population level, the underlying
22 neural and cognitive processes by which individuals incorporate information from connected peers
23 remain to be explored. Answers to this question would shed light on the mechanisms of social learning
24 and decision-making in a wider and more ecologically-relevant context, and help understand collective
25 maladaptation—such as herding⁹ and misinformation propagation¹⁰—in terms of the computational
26 challenges faced by individuals trying to process entangled information in an interconnected society.

27

28 A natural candidate for investigating learning on social networks from a neurocognitive perspective is
29 the computational framework of reinforcement learning (RL)¹¹. RL theories have been highly successful
30 in connecting the cognitive and neurobiological bases with a broad range of behaviors¹²⁻¹⁴, including
31 learning from social partners¹⁵⁻²⁰. Despite its successes, the standard RL framework provides an
32 incomplete account of learning in complex, interconnected environments. Consider, for example, the
33 classic observational learning task²⁰⁻²³, where an individual learns from the decisions of multiple
34 observees performing the same task as the observer. Prior RL-based research has typically assumed that
35 the actions of different observees constitute independent learning signals, which can be statistically
36 aggregated as an unbiased estimator for the common, unknown state of the environment for the
37 observer^{20,23,24}. Contrary to this assumption, however, substantial evidence in the social network
38 literature suggests that choices selected by connected peers are often interrelated and vary in their
39 informativeness^{25,26}. Blindly relying on the conventional RL strategy without considering the
40 underlying connections that spread social influences is essentially ignoring the potential variations and
41 repetitions in social signals and can be detrimental to adaptive behavior in an interconnected
42 environment²⁷. Nevertheless, extant data suggest that social animals embedded in complex interaction
43 webs demonstrate some level of sensitivity to the topological features of their immediate social
44 environments^{28,29}, leaving open whether and how relevant structural information is incorporated into
45 the learning processes.

46

47 Theoretically, the social learning literature has proposed two classes of models to address the network
48 effect. Normative strategies, such as Bayesian learning, assume that individuals rationally use the
49 knowledge of the network structure to optimally distinguish between learning signals, filter out
50 potential correlations in those signals, and integrate only the new piece of information into belief with
51 the Bayes' rule^{7,30}. Despite its theoretical appeal, converging evidence suggests that Bayesian learning
52 is cognitively unrealistic due to its excessive computational demand even on networks with relatively
53 simple structures³⁰. Naïve strategies, such as DeGroot learning, take an opposite, heuristic approach³¹.
54 Although these models do not optimally adjust for signal heterogeneity and interdependency and
55 sometimes lead to incorrect consensus among network members^{7,30}, they provide simple quantitative
56 accounts for how network geometry may affect learning. For instance, the now canonical DeGroot
57 learning theory posits that learning on a network can be approximated by a Markov process driven by
58 a weighted average of signals received from network neighbors. The weight used for signal aggregation
59 reflects how strongly a particular individual is influenced by a neighbor, and has been linked
60 theoretically to the network structure based on the limiting property of Markov processes³². However,
61 no direct evidence is available for the bounded rational assumption at the heart of the naïve learning
62 theories or for whether and how the underlying neurocognitive operations related to social influence

63 are affected by network structures in ways that translate to trial-by-trial (rather than asymptotic)
64 behavior^{30,33}.

65

66 We hypothesized that the process by which the brain learns from networked environments could be
67 characterized by incorporating the DeGroot heuristics into the RL framework. In the context of learning
68 observationally from others' decisions, the proposed DeGroot-RL model stands on the following three
69 hypotheses that together enable us to delineate the rich network interactions in a quantitative yet
70 neurobiologically-plausible manner. First, similar to the previous temporal difference algorithms of RL
71 widely used in nonnetworked settings¹¹, the DeGroot-RL model assumes that the brain integrates
72 information across social observations by maintaining and updating an internal expectation signal.
73 Learning from an action of a particular neighbor is driven by an action prediction error (aPE) between
74 the observed and expected choice of this neighbor, weighted by a learning rate.

75

76 Second, the networked environment affects learning by differentially modulating the learning rate.
77 Motivated by prior data on social influence that a well-connected individual has greater influence on
78 her peers and is less susceptible to others' opinion³⁴, the DeGroot-RL model posits that the extent to
79 which one learns from an observed action scales with the observee's network connectedness, relative
80 to that of the observer. Under this assumption, the brain needs to flexibly adjust the learning rate based
81 on network locations, possibly according to signals related to the observee's and observer's degree
82 centralities (i.e., the number of individuals to whom one is directly connected on a network), one of the
83 most fundamental metrics for local prominence and immediate influence in social network analysis².

84

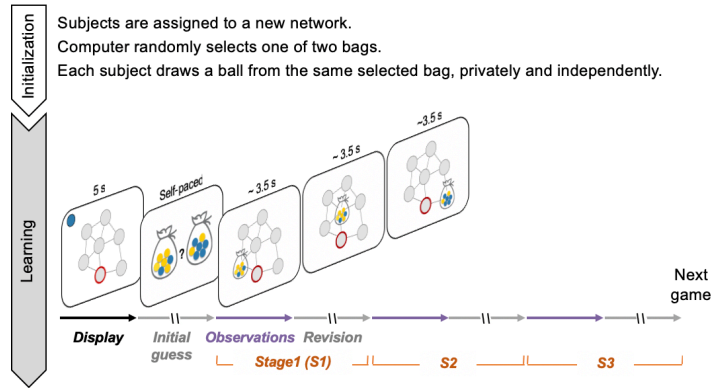
85 The third hypothesis, derived from the DeGroot heuristics, postulates that the degree-modulation effect
86 on learning may vary systematically over the course of information circulation, contingent on whether
87 social observations differ in their informativeness³⁰. For example, when individuals learn from others'
88 firsthand, isolated information, the DeGroot-RL model will reduce to the standard RL-like algorithm
89 for observational learning^{21,22}, whereby an observer is equally influenced by the received information
90 regardless of the differences in observees' network locations or properties. In contrast, when learning
91 from others' secondhand, possibly heterogeneous and intertwining information, the strength of learning
92 will be modulated by the relative degree centrality between the observee and the observer on the
93 network. Under this hypothesis, the brain needs to process network-related information flexibly,
94 according to its relevance to learning.

95

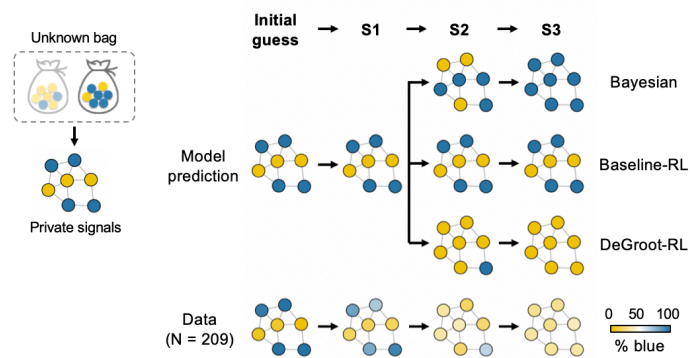
96 To assess these hypotheses and elucidate the neural and cognitive process related to the proposed
97 DeGroot-RL algorithm, we used functional magnetic resonance imaging (fMRI) in conjunction with a
98 distributed learning task for observational learning^{30,35,36} that was adapted from economic studies of
99 information cascade³⁷ and housed in a variety of exogenously given, 7-node, undirected, and
100 unweighted networks. This task is simple enough to carry out in a controlled fMRI experiment with a
101 reasonably large number of individuals interacting with one another in real time, yet it retains important
102 features of opinion adaptation under social influence. Moreover, rather than focusing on a handful of
103 special networks as in prior behavioral experiments^{3,30}, we investigated the proposed model on a
104 relatively large variety of network structures that were preselected based on the separability of choice
105 behavior simulated by different learning models (*Methods*).

106

A Task schematic



B Example



C DeGroot-RL model

S1:

$$E_{new}(\text{bag}) = E_{old}(\text{bag}) + \alpha \cdot aPE \rightarrow 1 - E_{old}(\text{bag})$$

S2 & S3:

$$E_{new}(\text{bag}) = E_{old}(\text{bag}) + \beta \cdot RD \cdot aPE$$

ND
Total local degree

107

108

109 **Fig. 1 | Distributed learning game and the DeGroot-RL model. (A)** Task schematic. Each distributed learning game is

110 initialized by randomly assigning subjects to different nodes on a new network and selecting one of two bags that contain

111 yellow and blue balls with opposing ratios (5:2 vs. 2:5). Each subject, who does not know which bag was selected, privately

112 draws a ball from the same selected bag with replacement, and needs to infer whether the underlying bag contains more blue

113 or yellow balls in a series of decisions. All subjects are instructed that the chance of drawing a blue/yellow ball is either $\frac{5}{7}$ or

114 $\frac{2}{7}$, independently and identically distributed across all members embedded on the same network. After initialization, subjects

115 are presented with the structure of the network (common for all network members; *Display*), one's own network location (red

116 circle in *Display*), and the ball privately drawn from the bag (top left corner in *Display*). Subjects are then simultaneously

117 asked to decide between two candidate bags (*Initial guess*), followed by three stages of observational learning (*S1–S3*). At

118 each stage, a participant is presented with the most recent decisions by her immediate neighbors (*Observations*) and then

119 provided with an opportunity to reassess her previous decision (*Revision*). To facilitate the visual tracking of neighbors' choices

120 on a network display while counterbalancing the influence of observation order, neighbors' choices are revealed sequentially,

121 in a clockwise order, starting from a randomly-selected neighbor that varies across stages and between subjects. All 7

122 participants played the same game in real time from their respective network locations, with no feedback on the choice

123 accuracy during the experiment (*Methods* and Supplementary Video 1). **(B)** Learning dynamics on an example network, which

124 also serves to illustrate possible misinformation propagation on the network. Left: the underlying bag selected by the computer

125 represents a simulated choice (model prediction) or the frequency of actual choices (data) over bags containing more blue or
126 yellow balls. While the Bayesian learning, baseline-RL, and DeGroot-RL models make similar predictions in the initial guess
127 and S1 estimation, they differ sharply in S2 and S3 predictions, with the DeGroot-RL predictions best aligned with the actual
128 choice frequencies of all participants on this network. Notably, this example also demonstrates a scenario in which network
129 members reach an incorrect consensus under the DeGroot-RL strategy. Unlike rational Bayesian learners, who gradually form
130 a consensus on the correct underlying state, both the simulated DeGroot-RL agents and actual participants converged toward
131 the wrong estimation, biased by the inaccurate information from the central, most-connected individual on the network (see
132 also Extended Data Fig. 4 for model simulation on all networks in the study). (C) DeGroot-RL model illustration. Left: An
133 example illustrating the stage-varying information in an observed action. While an S1 observation reflects the neighbor's
134 private signal, an observation in S2 or S3 additionally signals what the neighbor has previously learnt from her neighbors.
135 Right: Stage-dependent, degree-modulated learning. Upon observing an action from a neighbor, the belief expectation about
136 the unknown state (E_{old}) is updated through an action prediction error (aPE), defined as the discrepancy between the observed
137 and expected action. Learning in S1 follows the typical RL setup where the aPE signal is scaled by a baseline learning rate
138 (α). In S2 and S3, however, aPE signals are weighted by the learning rate (β) and relative degree (RD), with the latter being
139 defined as the degree centrality of the observee relative to the total degree of the observer and all her direct neighbors on the
140 network (*Methods*). Variants of DeGroot-RL formulations such as alternative learning rate specifications and RD definitions
141 were evaluated against the proposed model in *Methods*.
142

143 Results

144 **A distributed learning game.** A total of 217 unique subjects (31 fMRI participants) participated in the
145 experiment in groups of 7 (1 inside the fMRI scanner; 209 included in data analyses with 25 fMRI
146 participants; see *Methods* for subject exclusion). The experiment consisted of 40 separate games on
147 varying networks (Extended Data Figs. 2-3; *Methods*). In each game, a participant's goal was to infer
148 an unknown state of the environment, which was common to all 7 participants in the game. At the
149 beginning of a game, 7 subjects were randomly assigned to different nodes on a new network, and a
150 computer selected one of two underlying states at random. Each subject received a private signal that
151 was independently and identically distributed conditional on the same selected state, and needed to
152 make an initial guess about the underlying state (Fig. 1A and Extended Data Fig. 1; *Methods*). We
153 hypothesized that a subject should rationally base her estimation on the private signal in this decision.
154 The prediction was confirmed by our data, in which $98.34 \pm 5.12\%$ (mean \pm intersubject SD) initial
155 estimations matched subjects' private signals.
156

157 Critical for the purpose of this study, the participants were then allowed to revise their estimations in
158 response to the choices previously selected by the neighbors to whom each was directly connected on
159 the network. To allow for meaningful fMRI analyses, a subject was presented with her direct neighbors'
160 prior decisions sequentially, one at a time, such that her neural responses could be directly linked to the
161 action and network location of a particular observee (Fig. 1A and Extended Data Fig. 1; *Methods*). To
162 allow for examining learning effects, the process of observing neighbors' actions and reassessing one's
163 estimation was carried out 3 times consecutively within each game (henceforth *Stages 1-3*, or *S1*, *S2*,
164 and *S3*; Fig. 1A).
165

166 Of note, 7 participants played the same game simultaneously, from their respective network locations
167 via an intranet. That is, the participants were facing the same underlying state, the same network
168 structure and display, and were making decisions at the same time in each game. Crucially, when a
169 participant was witnessing her neighbors' choices, her neighbors were also presented with the choice
170 information from their respective neighborhoods (see illustration in Supplementary Video 1). Under
171 such a real-time distributed learning setup, information received by an observer is incorporated into the
172 observer's subsequent decision and propagated gradually from the observer to her direct and indirect
173 contacts along network connections in the later stages of the game. Throughout the experiment, all

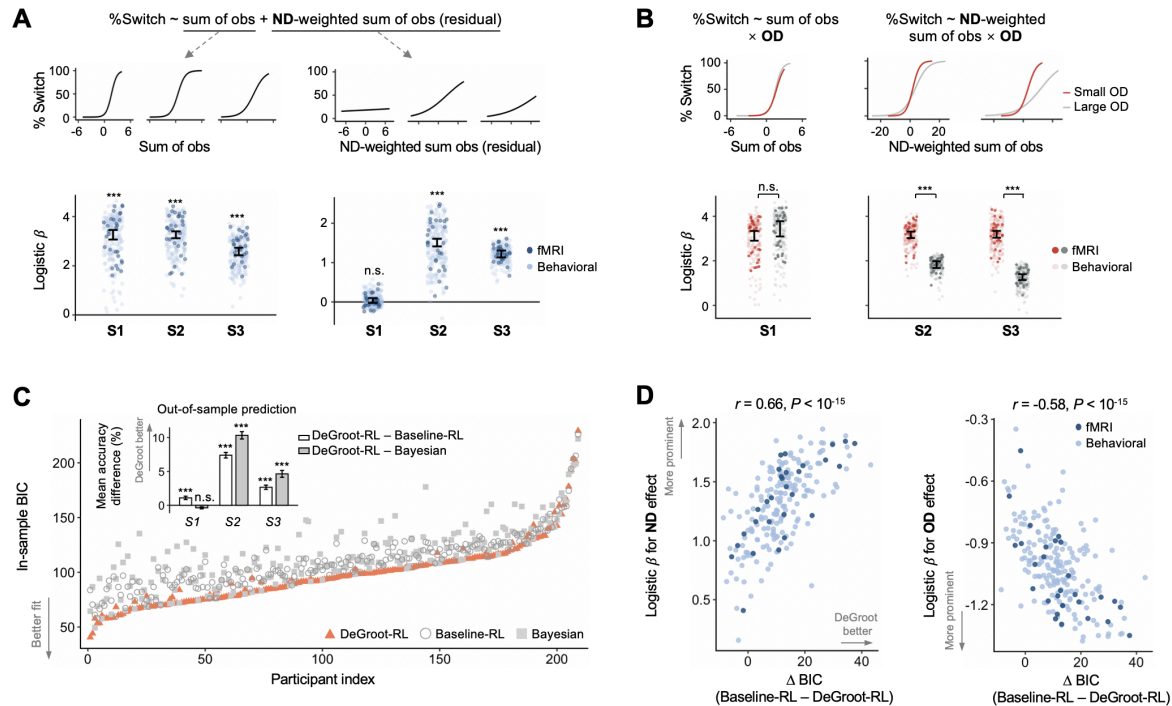
174 subjects were financially incentivized to guess as accurately as possible in all 4 decisions in each game
175 (i.e., initial guess and 3 reassessments) and had no incentive to mislead or collude with others. No
176 feedback was provided on the accuracy of estimations during the experiment, so that the only
177 information a subject could rely on in a game was her private signal and the actions by her direct
178 neighbors on the network (*Methods*).

179
180 This real-time, distributed learning game has two properties important for evaluating the DeGroot-RL
181 model at the neurocomputational level. First, the three learning stages (S1–S3) were set up identically
182 in each game (but with randomized observation order to control for potential order effects; *Methods*).
183 This feature allowed us to evaluate the first DeGroot-RL hypothesis that subjects learned from
184 neighbors' actions in an error-driven manner and did so consistently across S1, S2, and S3. Second,
185 despite their identical experimental setups, three learning stages differed in the type of information
186 contained in neighbors' actions. Unlike S1 observations, which reflect observees' independent private
187 signals, an S2 or S3 observation would additionally signal what the observee has learned from her
188 neighbors (Fig. 1C, left), thereby becoming relatively more informative to the specific observer when
189 the observee is better-connected and the observer is less-connected. Given this feature, the last two
190 DeGroot-RL hypotheses led to a precise and testable prediction for the experiment—that is, learning
191 from network neighbors might be modulated by the degree centrality of the observee relative to the
192 observer in S2 and S3, but not in S1.

193
194 According to these predictions, we should expect not only the stage-dependent degree-modulation
195 effect at the behavioral level, but also the differential involvement of neural signals of aPE and relative
196 degree centrality—two core computational components in the DeGroot-RL model (Fig. 1C, right)—in
197 the learning process. We predicted that there should be brain regions that track the aPE-related signals
198 nonselectively in S1, S2, and S3, as well as brain regions that respond to signals associated with the
199 relative degree centrality in S2 and S3, but not in S1.

200
201 **Learning behavior was modulated by degree centralities in S2 and S3, but not in S1.** Behaviorally,
202 participants in the experiment adapted their decisions in response to neighbors' choices, such that the
203 level of consensus within a network grew from $61.21 \pm 1.29\%$ (mean \pm intergroup SD) in the initial
204 guess to $88.10 \pm 3.03\%$ in the last (S3) decision, with no significant difference in the level of consensus
205 changes across games between the early and late parts of the experiment (Pearson's correlation between
206 consensus change and game order: $r = -0.05$, $P = 0.093$; see also Extended Data Fig. 6 for behavioral
207 dynamics in choice accuracy and more). To characterize the overall learning effect and evaluate
208 DeGroot-RL predictions, we first performed model-free logistic regression analyses in each separate
209 stage, to examine the extent to which the likelihood of participants altering their choices was influenced
210 by social observations, and whether the strength of influence was modulated by the degree centralities
211 of the observees and observers. We hypothesized that the likelihood that a participant aligned her
212 estimation to an observation would be positively associated with the neighbor's degree centrality (ND)
213 but negatively associated with the observer's own degree (OD) in S2 and S3, but not in S1.

214



215
 216 **Fig. 2 | Behavioral evidence supporting the DeGroot-RL model.** (A) Neighbor’s degree centrality (ND) modulates learning
 217 in S2 and S3, but not in S1. Logistic regression tests the probability of an observer modifying her previous estimation against
 218 the (unweighted) sum of observations (left) and the weighted sum of observations, with each respective neighbor’s degree
 219 serving as the weight (right). The ND-weighted regressor was orthogonalized against the unweighted regressor to remove
 220 shared variances. Top: the probability of changing a decision as predicted by the logistic regression estimates at each stage.
 221 Bottom: fixed-effects regression coefficients and SEM for the unweighted and ND-weighted sum of observations (depicted by
 222 error bars), overlaid by the random-effect coefficient associated with each observer (depicted by dots). Each dark/light dot
 223 represents an fMRI/behavioral participant. (B) Observer’s own degree centrality (OD) is negatively associated with the
 224 susceptibility to social observations in S2 and S3, but not in S1. The top and bottom panels show results from logistic regression
 225 testing the likelihood of changing one’s estimation in response to social observations, when an observer is highly vs. poorly
 226 connected. For the illustration purpose, large and small ODs were defined by median splits on OD across networks for each
 227 participant. (C) Comparisons of in-sample model fits using the Bayesian Information Criterion (BIC) of each participant show
 228 that DeGroot-RL explains behavioral choices better than Bayesian and baseline-RL (paired *t*-test, DeGroot-RL vs. baseline-
 229 RL, mean ± SEM = -11.58 ± 0.68, $t_{208} = -17.07, P < 10^{-15}$; DeGroot-RL vs. Bayesian, -16.02 ± 1.14, $t_{208} = -14.05, P < 10^{-15}$;
 230 see also Extended Data Fig. 5 for Bayesian model selection). Participants are sorted by the BIC score of the model that best
 231 explains choices. Inset: Out-of-sample prediction accuracy is superior for the DeGroot-RL model compared with Bayesian
 232 and baseline-RL models in S2 and S3. The observed difference in the prediction accuracy in S1 between DeGroot-RL and
 233 baseline-RL likely reflects the fact that the model parameters governing S1 predictions (e.g., α and inverse temperature)
 234 were set to maximize the likelihood of choices across all stages (rather than S1 only) and thus were affected by model configuration
 235 and estimation in other stages. (D) Across-subject correlation between model-free and model-based results. Individual BIC
 236 differences between DeGroot-RL and baseline-RL (*x*-axis) are plotted against the individual model-free estimates for ND (left)
 237 and OD (right) effects, respectively, exploiting the fact that the DeGroot-RL model differs from the baseline-RL model only
 238 in the assumption regarding the degree-modulation effect in S2 and S3 (see model setup in Fig. 1C and *Methods*). The model-
 239 free ND effect (*y*-axis, left) is captured by the coefficient of individual random effects with respect to the ND-weighted sum
 240 of observations (as shown by the dot in the bottom-right panel in Fig. 2A), averaged across S2 and S3 within each subject.
 241 The OD effect (*y*-axis, right) reflects the coefficient of individual random effects for OD × ND-weighted sum observations (as
 242 shown by the within-subject difference between the orange and grey dots in the bottom-right panel in Fig. 2B), averaged across
 243 S2 and S3 within each subject. A more negative OD effect (*y*-axis, right) corresponds to a more prominent reduction in the
 244 susceptibility to social influence as OD increases. These correlations remained significant when analyzing the across-subject
 245 association for S2 and S3 separately (Pearson’s correlation between Δ BIC and ND effect in S2, $r = 0.62, P < 10^{-15}$, in S3, $r =$
 246 $0.32, P < 10^{-5}$; correlation between Δ BIC and OD effect in S2, $r = -0.62, P < 10^{-15}$, in S3: $r = -0.14, P = 0.039$). Consistent
 247 results were also found after partialing out the variance shared between the model-free OD and ND estimates (partial

248 correlation between Δ BIC and ND effect, $\rho = 0.44$, $P < 10^{-10}$; Δ BIC and OD effect, $\rho = -0.25$, $P < 0.001$). Error bars represent
249 SEM. *** $P < 0.001$, n.s., not significant, Bonferroni-corrected when appropriate.

250

251 To assess the impact of ND, we summarized the sequence of observations revealed to a subject at each
252 stage of each game, using two variables: (i) sum of observations across all direct neighbors (1, if an
253 observed action differs from the observer's previous choice; otherwise, -1), and (ii) weighted sum of
254 observations across all direct neighbors, with the respective neighbor's degree serving as the weight.
255 The inclusion of the unweighted and ND-weighted regressors helped to isolate the degree-modulation
256 effect of interest from a baseline tendency of following the majority, a phenomenon widely reported in
257 studies of group decision-making^{9,38}. To remove any shared variances between the two regressors, we
258 orthogonalized the ND-weighted regressor against the unweighted regressor, such that choices that were
259 equally explainable by the two variables were attributed solely to the unweighted regressor. The
260 regression coefficient for the ND-weighted sum of observations, therefore, served as a more stringent
261 test on whether ND modulated learning.

262

263 Consistent with our hypothesis, mixed-effects regression conducted separately for each learning stage
264 showed that the likelihood of a participant modifying her decision was positively correlated with not
265 only the unweighted sum of observations, but also the ND-weighted sum of observations, in both S2
266 and S3 (Fig. 2A; sum of obs. in S2: $\beta \pm \text{SEM} = 3.26 \pm 0.14$, $z = 23.40$, $P < 10^{-15}$; sum of obs. in S3, β
267 $= 2.58 \pm 0.15$, $z = 16.91$, $P < 10^{-15}$; ND-weighted sum in S2: $\beta = 1.51 \pm 0.10$, $z = 15.68$, $P < 10^{-15}$; ND-
268 weighted sum in S3: $\beta = 1.22 \pm 0.09$, $z = 14.10$, $P < 10^{-15}$). The positive effects suggested that, in
269 addition to following the majority, subjects were more likely to be swayed toward the decisions of
270 highly connected neighbors, relative to those of poorly connected neighbors embedded on the same
271 network. In stark contrast, in S1, the ND-weighted regressor showed no extra explanatory power above
272 and beyond the unweighted regressor in predicting participants' subsequent choices (Fig. 2A; sum of
273 obs. in S1, $\beta = 3.26 \pm 0.19$, $z = 16.81$, $P < 10^{-15}$; ND-weighted sum in S1, $\beta = 0.04 \pm 0.06$, $z = 0.63$, $P =$
274 0.528 ; see also Supplementary Table 2A for full regression results).

275

276 To assess how learning was biased by a participant's OD, we further compared the influence of the ND-
277 weighted sum of observations on the observer's subsequent decision when the observer was endowed
278 with high vs. low degree centrality. Specifically, we carried out separate mixed-effects logistic
279 regression for S2 and S3, testing the probability of an observer modifying her decision against the
280 following variables: OD, ND-weighted sum of observations, and the interaction between these two
281 variables. These analyses showed positive main effects for both the OD and ND-weighted regressors
282 (Supplementary Table 2C). More importantly, and consistent with the DeGroot-RL prediction, a
283 negative interaction was seen between these variables in both S2 and S3 (Fig. 2B; interaction in S2: β
284 $= -1.10 \pm 0.11$, $z = -9.69$, $P < 10^{-15}$; S3: $\beta = -1.00 \pm 0.09$, $z = -11.43$, $P < 10^{-15}$), suggesting a decreased
285 susceptibility to social observations when a participant was highly relative to poorly connected. This
286 result was not due to the specific way of summarizing the sequence of observations within each stage
287 and remained significant when examining the interaction between OD and the unweighted (rather than
288 ND-weighted) sum of observations (Supplementary Table 2B). In S1, by comparison, no systematic
289 variation was observed in participants' susceptibility to neighbors' actions with their OD, as revealed
290 by a similar regression analysis for the likelihood of altering one's decision against variables of OD,
291 unweighted sum of observations, and the interaction between the two (Fig. 2B; interaction in S1: $\beta =$
292 0.15 ± 0.17 , $z = 0.92$, $P = 0.357$; see also Supplementary Table 2B).

293

294 **The DeGroot-RL model characterized learning behavior better than alternative models.** To more
295 formally test the DeGroot-RL predictions and derive latent variables that might reflect neurocognitive
296 operations underlying learning on networks, we fit the DeGroot-RL model with each participant's
297 choice behavior (*Methods*). As shown in Fig. 1C, our proposed DeGroot-RL model assumed that the
298 S2 and S3 learning rate was modulated by a measure of relative degree (RD), defined as the observee's
299 degree with respect to the total degree of the observer and all her direct neighbors on the network
300 (*Methods*). The normalization term in RD reflected the mathematical requirement that a learning rate
301 should be no greater than 1, although the alternative specification using the observer's degree as the
302 denominator yielded similar results (*Methods* and Supplementary Table 1).

303

304 We compared the proposed DeGroot-RL model against two benchmark models: a baseline-RL model,
305 which assumed that the network-related information was completely ignored throughout learning, and
306 a Bayesian learning model, which assumed that the information regarding the network structure was
307 rationally used by all participants (*Methods*). It is worth noting that the DeGroot-RL model differed
308 from the Bayesian and baseline-RL models in predicting choices mainly in S2 and S3, but was similar
309 to those models in S1, such that all three models predicted the unweighted, rational integration of S1
310 observations³⁰ (*Methods*; see also Fig. 1B for an example).

311

312 We computed both the Bayesian Information Criterion (BIC) based on the in-sample model fit and out-
313 of-sample prediction accuracies using a five-fold cross-validation procedure (*Methods*; see also
314 Supplementary Table 3 for model estimation results). Both measures showed that the DeGroot-RL
315 model outperformed the alternative models. In particular, the DeGroot-RL model had a lower BIC
316 (better fit) than the alternative models (Fig. 2C; DeGroot-RL = 20407.41; baseline-RL = 22827.97;
317 Bayesian = 23755.10), with its BIC score being the lowest in more than 78% of participants (164 out
318 of 209 participants; see also Extended Data Fig. 5 for Bayesian model selection). Based on the
319 prediction accuracy computed from holdout samples, the DeGroot-RL model predicted subjects'
320 behavior with an accuracy of $75.29 \pm 0.62\%$ (mean \pm intersubject SEM, averaged over all four decisions
321 within each game), which was significantly higher than that from the baseline-RL model (paired *t*-test,
322 $t_{208} = 15.38$, $P < 10^{-15}$) and the Bayesian learning model ($t_{208} = 15.87$, $P < 10^{-15}$; see also Extended Data
323 Fig. 5). This increase in the prediction accuracy was largely contributed by the improved prediction in
324 both S2 and S3, rather than an enhancement coming from either S2 or S3 alone (inset, Fig. 2C).

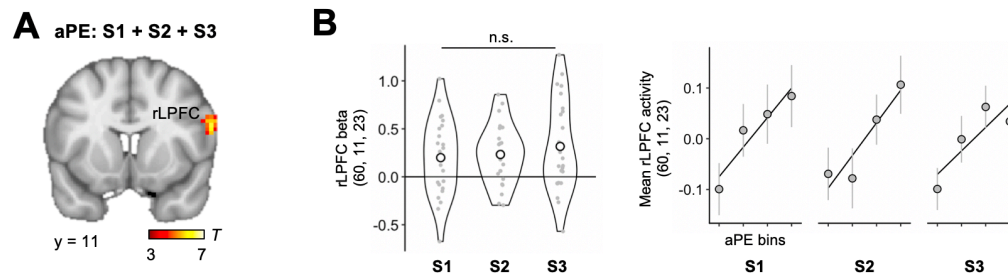
325

326 Importantly, choice simulation based on the DeGroot-RL model estimates could successfully recover
327 the model-free patterns in each learning stage (Extended Data Fig. 7). In contrast, simulations from the
328 Bayesian and baseline-RL models—even with the best-fitting parameters calibrated on choice
329 behavior—failed to capture key behavioral features in S2 and S3, such as the ND and OD effects
330 revealed by the logistic regression analyses of the actual data (Extended Data Fig. 7). Besides the
331 aggregate choice patterns, the DeGroot-RL model estimates were also consistent with the model-free
332 analyses at the across-subject level, such that subjects whose behavior was better characterized by
333 DeGroot-RL showed a more pronounced behavioral sensitivity to both ND and OD in S2 and S3, as
334 measured by the respective logistic regression estimates for individual participants (Fig. 2D).

335

336 Besides individual choices, the proposed DeGroot-RL model also captured the variations in other
337 behavioral measures, including choice difficulty as reflected by participants' reaction time, the
338 dynamics of choice consensus, and the trajectory of estimation accuracy (Extended Data Fig. 6). In
339 addition to the Bayesian and baseline-RL models, the proposed model also outperformed a range of

340 alternative models that could account for participants' choices based on assumptions differing in either
341 which network measure (rather than RD) might modulate learning or how RL parameters were specified
342 in the model (*Methods* and Supplementary Table 1). Finally, to address the potential concern regarding
343 the between-network variations in learning behavior, we conducted model estimation for each network
344 (pooling over subjects), and observed similar results using both the in-sample and out-of-sample
345 measures for goodness-of-fit at the across-network level (*Methods* and Extended Data Fig. 5).
346



347
348 **Fig. 3 | Right lateral prefrontal cortex (rLPFC) tracks the value estimate of action prediction error (aPE) in S1, S2, and**
349 **S3. (A)** Statistical parametric map with respect to aPE estimates at observation onsets, computed by averaging the GLM1
350 coefficients for aPE estimates across S1, S2, and S3 for each subject and then taking them into the standard random-effects
351 group analysis (peak voxel Montreal Neurological Institute (MNI) coordinates: $x, y, z = 60, 11, 23$; cluster-wise family-wise-
352 error(FWE)-corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$). **(B)** rLPFC cluster demonstrates similar effect
353 sizes for individual aPE estimates across stages. Left: Violin plots for the distribution of individual rLPFC beta values for aPE
354 estimates, separately extracted for each stage from the significant rLPFC cluster as identified in Fig. 3A. Each dot represents
355 a subject. Each circle represents a group mean. Right: Mean fMRI activity extracted from the same rLPFC cluster and binned
356 by aPE estimates in each separate stage. Similar results were obtained by a whole-brain paired comparison, showing that no
357 significant cluster differentially responded to aPE estimates across stages at cluster-wise FWE-corrected $P < 0.05$. N.s., not
358 significant. Error bars represent intersubject SEM.

359
360 **Right lateral prefrontal cortex (rLPFC) tracked aPE estimates in S1, S2, and S3.** Having
361 established the DeGroot-RL model at the behavioral level, we then investigated whether fMRI activity
362 reflected key computational components of the model, including aPE and RD, on an observation-by-
363 observation basis, at the time when participants were presented with a neighbor's action. We also tested
364 whether the neural responses to the aPE and RD signals would demonstrate differential stage-related
365 patterns, as predicted by the DeGroot-RL model. We conducted a standard general linear model (GLM)
366 analysis on fMRI data, entering the value estimate of aPE associated with each observed action derived
367 from the best-fitting DeGroot-RL model for each individual, together with the RD value associated with
368 the respective observee and observer, at the observation onsets of the corresponding learning stage
369 (GLM1, *Methods*). Parametric regressors were orthogonalized against one another in all GLM analyses
370 in the present study, such that the regression coefficients captured the variations in blood-oxygen-level-
371 dependent (BOLD) signals in the specific brain regions that were uniquely explained by each regressor,
372 rather than the shared variances.

373
374 As we were particularly interested in evaluating whether there were any brain regions consistently
375 tracking aPE estimates across three learning stages, we averaged the GLM1 coefficients of the aPE
376 estimates across stages for each subject before taking them into the group-level analyses. This identified
377 significant neural responses in the rLPFC (Fig. 3A; see also Supplementary Table 5 for full activation
378 including the middle temporal gyrus and visual cortex), which has been previously implicated in
379 representing notions of prediction error signals in action observation learning^{21,22,24} and as part of the
380 "mirror" system encoding the executed and observed actions in a range of interpersonal scenarios³⁹.

381 Rather than being driven by a simple effect such as the correlation with the observed action, activity in
382 the identified rLPFC cluster reflected both computational components in an aPE signal: It scaled
383 positively with the observed action (1 if the observation matches the observer's previous decision; $\beta =$
384 0.23 ± 0.04 , $t_{24} = 5.29$, $P < 0.0001$; Extended Data Fig. 8), but negatively with the observer's expectation
385 about the underlying state ($\beta = -0.12 \pm 0.06$, $t_{24} = -2.23$, $P = 0.036$; Extended Data Fig. 8). As a
386 robustness check, we tested additional decision variables that might be related to the processing of
387 social observations, including color selected by the observee (yellow/blue), order in which the
388 neighbor's action was presented in the particular stage of the game, and the consensus level among
389 network members at the beginning of the stage (GLM2, *Methods*). The observed rLPFC encoding could
390 not be attributed to any of these variables and remained significant with the inclusion of these variables
391 as regressors of no interest in one regression model (cluster-wise family-wise-error(FWE)-corrected P
392 < 0.05 , with cluster-forming threshold uncorrected $P (P_{unc.}) < 0.001$).

393

394 Moreover, the rLPFC cluster stably tracked aPE estimates across S1, S2, and S3, demonstrating similar
395 effect sizes across stages in the regression estimates (beta values) that were extracted from the
396 significant rLPFC cluster for each separate stage (Fig. 3B, left; one-way repeated-measures ANOVA,
397 $F(2,48) = 0.55$, $P = 0.582$). To illustrate this result, we also plotted the mean activity obtained from the
398 rLPFC cluster as a function of four bins of ascending aPE estimates in separate learning stages and
399 observed similar correlation patterns (Fig. 3B, right). This result was further confirmed by a whole-
400 brain within-subject comparison of the aPE correlates, in which we identified no significant cluster that
401 responded differentially to aPE estimates across stages at cluster-wise FWE-corrected $P < 0.05$.

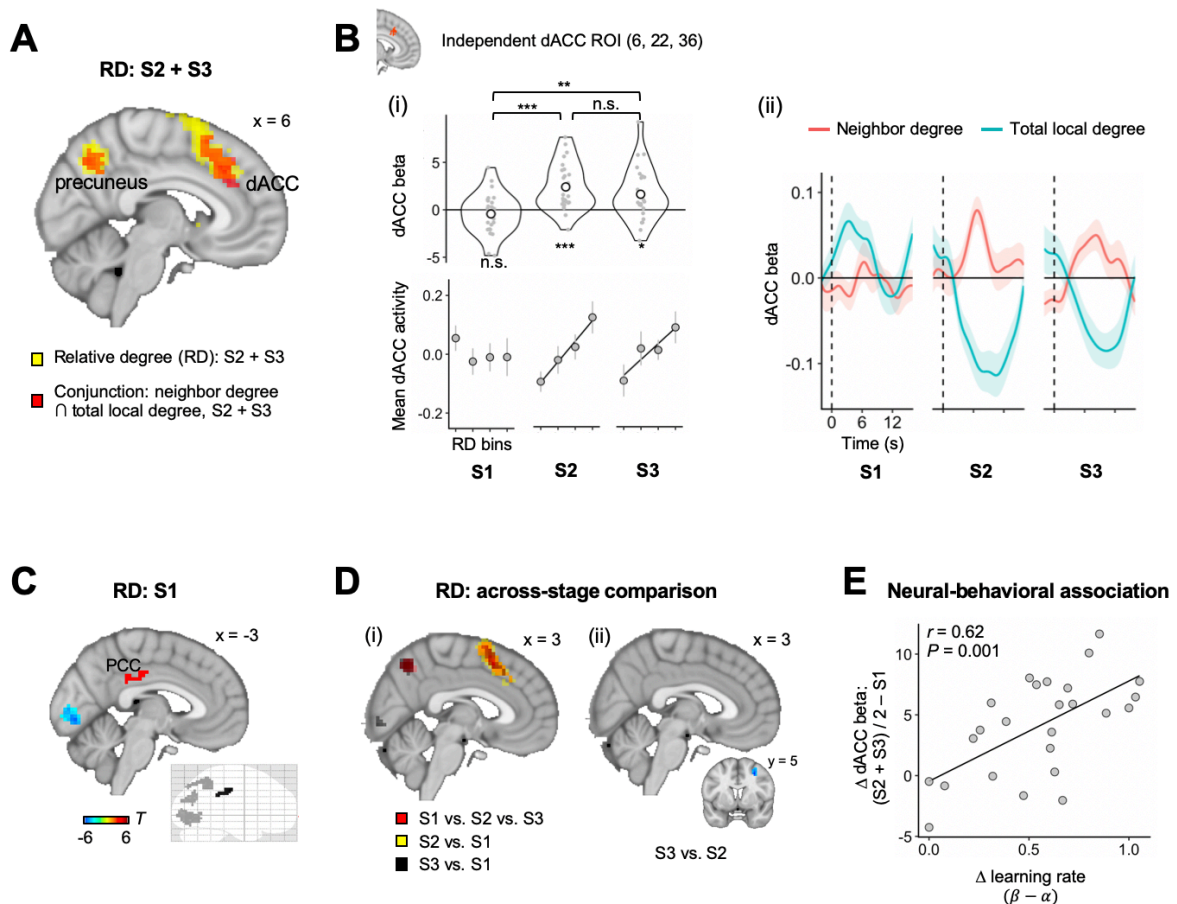
402

403 In addition to aPE signals at observation onsets, we also tested for brain regions responding to other
404 classic learning signals at the time of decision submission in each learning stage. Consistent with prior
405 neuroimaging evidence, we observed activity in the orbitofrontal cortex (OFC) signaling the reward
406 expectation estimate associated with the chosen option^{13,14}, as well as activity in the anterior cingulate
407 cortex (ACC) and the neighboring medial prefrontal cortex reflecting the model-derived probability of
408 switching away from one's prior estimation⁴⁰ (Extended Data Fig. 9). Findings at the choice time,
409 together with the aPE signals at the observation time, indicated an error-driven process similar to the
410 temporal difference form of RL during action observation learning channeled by social networks.

411

412 **Dorsal anterior cingulate cortex (dACC) represented RD-related signals in S2 and S3.** Next, we
413 tested whether activity in any brain regions would reflect the observation-by-observation changes in the
414 relative degree centrality between the observee and the observer (i.e., RD) at observation onsets, and
415 did so stably across S2 and S3. Similar to the analysis of aPE signals, we first looked for the RD
416 correlates by averaging the GLM1 coefficients for RD over S2 and S3 for each fMRI participant
417 (*Methods*). This revealed a strong correlation in a network of brain regions, including the dorsal anterior
418 cingulate cortex (dACC) extending to the adjacent presupplementary motor area (preSMA), precuneus,
419 bilateral anterior insula, visual cortex, and other areas (Fig. 4A; see also Supplementary Table 5 for full
420 activation list). The loci of activation in the dACC/preSMA were similar to those seen in the past
421 experiments where subjects adjusted behavioral strategies, such as learning rate, in response to
422 environmental changes⁴¹⁻⁴⁵. Moreover, dACC activity demonstrated features consistent with the
423 assumption that the relative rather than absolute value of degree centrality was involved in learning.
424 Activity in the dACC simultaneously correlated with the neighbor's degree (numerator in RD) and total
425 local degree (denominator in RD), with opposing signs, at observation onsets in both S2 and S3.
426 Specifically, this opposing correlation pattern was observed not only within a region of interest (ROI)

427 in the dACC independently defined using an automated online meta-analysis⁴⁶ (Fig. 4B, right), but also
 428 in a whole-brain conjunction analysis across positive activation for neighbor's degree and negative
 429 activation for the total local degree in S2 and S3 (Fig. 4A).
 430



431
 432 **Fig. 4 | Activity in the dorsal anterior cingulate cortex (dACC) correlates with the value of relative degree (RD) in S2**
 433 **and S3, but not in S1. (A)** dACC shows RD-related signals on an observation-by-observation basis in S2 and S3. Regions
 434 shaded in yellow indicate clusters where activity significantly correlates with RD values at observation onsets in S2 and S3,
 435 calculated by averaging the GLM1 coefficients for RD across S2 and S3 for each participant and then taking them into the
 436 standard group-level analysis (cluster-wise FWE-corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$; see also
 437 Extended Data Fig. 10A for the extent of RD encoding). Regions shaded in red indicate clusters scaling both positively with
 438 neighbor's degree (numerator in RD) and negatively with total local degree (denominator in RD) at observation onsets in S2
 439 and S3, as revealed by a whole-brain conjunction analysis for overlapping activation between the neighbor's degree and total
 440 local degree (cluster-wise FWE-corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$; *Methods*). The conjunction
 441 result was computed by first averaging the individual GLM coefficients for neighbor's degree (or total local degree; GLM4,
 442 *Methods*) across S2 and S3 in the same way we looked for the RD correlates, and then using the resulting statistical maps to
 443 test for overlapping activation between either (i) the positive encoding of neighbors' degree and negative encoding of total
 444 local degree (shown in Fig. 4A), or (ii) the negative encoding of neighbors' degree and positive encoding of total local degree
 445 (no significant overlap at cluster-wise FWE-corrected $P < 0.05$; *Methods*). **(B)** dACC region of interest (ROI), independently
 446 defined by Neurosynth⁴⁶. Top-left: beta values with respect to RD extracted from the same independent ROI at observation
 447 onsets in separate learning stages (S1: $\beta = -0.43 \pm 0.44$, $t_{24} = -0.99$, $P = 0.332$; S2: $\beta = 2.41 \pm 0.50$, $t_{24} = 4.85$, $P < 10^{-4}$; S3: β
 448 $= 1.64 \pm 0.56$, $t_{24} = 2.91$, $P = 0.008$; one-way repeated-measures ANOVA, $F(2,48) = 10.65$, $P < 0.001$). Bottom-left: mean
 449 dACC activity binned by RD values in each stage. Error bars represent intersubject SEM. Right: Time-course analyses with
 450 respect to the neighbor's degree and total local degree for each stage within the same independent dACC ROI. Vertical dashed
 451 lines indicate the observation onset. **(C)** Neural correlates of RD values at observation onsets in S1 (cluster-wise FWE-
 452 corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$; see also Extended Data Fig. 12 for ROI analyses in S1). **(D)**
 453 Paired comparisons with respect to RD correlates across learning stages. Left: Clusters shaded in red show results of whole-
 454 brain ANOVA analysis comparing RD correlates across S1, S2, and S3 within subjects. Clusters in yellow and black show

455 post hoc ANOVA analyses testing the stage effect by comparing the RD correlates in S2 vs. S1 (yellow) and S3 vs. S1 (black)
456 (all cluster-wise FWE-corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$). Right: Post hoc paired comparison of
457 RD correlates between S2 and S3. The only significant cluster locates in the middle frontal gyrus (MNI: $x, y, z = 30, 5, 47$;
458 cluster-wise FWE-corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$). (E) Across fMRI subjects, the dACC beta
459 values with respect to RD in S2 and S3, relative to that in S1, is positively correlated with the individually-estimated learning
460 rate in S2 and S3, relative to that in S1. This effect was significant not only when we averaged the individual's dACC beta
461 over S2 and S3 (Pearson's $r = 0.62$, $P = 0.001$), but also when we tested the effect in S2 and S3 separately, with no significant
462 difference between these stages (S2: $r = 0.52$, $P = 0.007$; S3: $r = 0.53$, $P = 0.006$; S2 vs. S3: $\beta = -0.11 \pm 0.25$, $t_{24} = 0.43$, $P =$
463 0.673). Each dot represents a subject. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, n.s.: not significant, Bonferroni-corrected when
464 appropriate.

465

466 Analyses further showed that the RD values explained dACC activity at observation onsets above and
467 beyond a range of decision-related variables, including visual properties associated with the network
468 display, nuisance effects arising from action observation, and cognitive components that might be
469 related to learning or to other dACC functions implicated by previous studies. In particular, we
470 performed another GLM analysis (GLM3, *Methods*), which additionally included the observee's visual
471 centrality in the network display (i.e., the Euclidean distance between the observee's location and the
472 visual center of the layout), the visual distance between the observee's and observer's location in the
473 network display, order of observation display, color selected by the observee (yellow/blue), magnitude
474 of aPE estimates, variance in attained observations within the current stage, level of conflict between
475 social observations and the observer's decision (i.e., the proportion of attained observations within the
476 current stage that were different from the observer's prior decision), updated belief expectation
477 associated with the observer's prior decision, and choice difficulty reflected by the distance in the belief
478 expectation estimates between two choice options. None of these variables could explain the same
479 portion of dACC activation as RD values and the observed parametric encoding of RD remained
480 significant in the dACC, even after regressing out the influence of all these variables as regressors of
481 no interest in the same GLM model (cluster-wise FWE-corrected $P < 0.05$, with cluster-forming
482 threshold $P_{unc.} < 0.001$; Extended Data Figs. 10-11).

483

484 Moreover, activity in the dACC demonstrated similar response patterns to the RD values in S2 and S3.
485 Neural betas separately extracted for S2 and S3 from the same independent dACC ROI were both highly
486 significant and showed no systematic difference in their effect sizes (Fig. 4B; S2: $\beta = 2.41 \pm 0.50$, $t_{24} =$
487 4.85 , $P < 10^{-4}$; S3: $\beta = 1.64 \pm 0.56$, $t_{24} = 2.91$, $P = 0.008$; S2 vs. S3: $\beta = 0.77 \pm 0.72$, $t_{24} = 1.07$, $P =$
488 0.294). This result was further confirmed by a whole-brain within-subject comparison of RD correlates.
489 Except for a cluster confined to the middle frontal gyrus (MNI: $x, y, z = 30, 5, 47$; Fig. 4D, right), we
490 found no other cluster that responded differently to RD values in S2 vs. S3 at cluster-wise FWE-
491 corrected $P < 0.05$ (Fig. 4D, right).

492

493 **Activity in the dACC did not correlate with RD values in S1.** By contrast, in S1, the same GLM1
494 analysis revealed no significant correlation with RD values at observation onsets in the dACC or other
495 frontal regions, in either positive or negative direction (Fig. 4C; cluster-wise FWE-corrected $P < 0.05$,
496 with cluster-forming threshold $P_{unc.} < 0.001$). Instead, we observed positive correlations with RD values
497 in a circumscribed cluster in the posterior cingulate cortex (PCC), and negative correlations restricted
498 to the precuneus and visual cortex (Fig. 4C and Extended Data Fig. 12). The identified regions in the
499 PCC and precuneus have been recently implicated in encoding features of real-world social networks,
500 even when such network features were task-irrelevant^{29,47,48}.

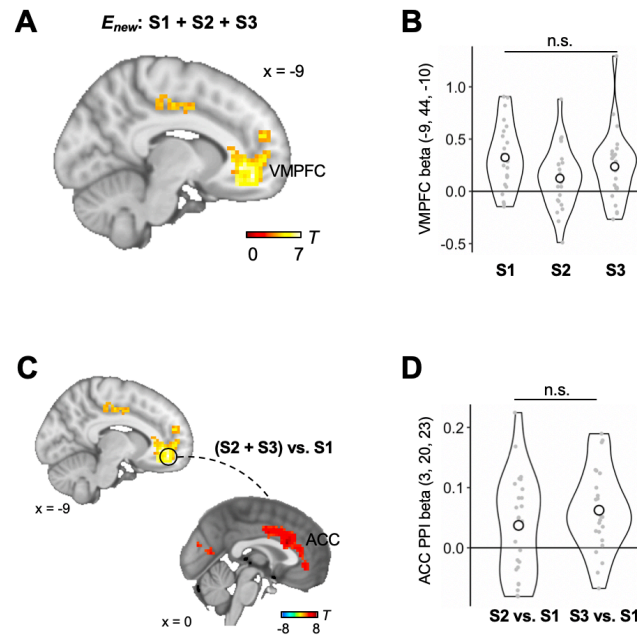
501

502 To more formally examine the spatial expression of the RD correlates and test the stage-varying
503 involvement of the dACC, we searched the whole brain for voxels that responded similarly (conjunction
504 analyses) or differently (ANOVA analyses) to RD values across stages. A three-way conjunction among
505 S1, S2, and S3 showed no significant cluster for either positive or negative activation to RD values
506 (cluster-wise FWE-corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$; Extended Data Fig.
507 10). Importantly, this lack of overlapping activation could not be attributed to the lack of overlaps
508 between S2 and S3, as an additional conjunction analysis between S2 and S3 identified substantial
509 activation to RD, including a large cluster in the dACC, that considerably overlapped with the activation
510 areas identified in Fig. 4A (Extended Data Fig. 10). Moreover, using a whole-brain ANOVA analysis,
511 we directly compared RD correlates across stages within subjects, and identified a significant stage
512 effect in several brain regions including the dACC (Fig. 4D, left; cluster-wise FWE-corrected $P < 0.05$,
513 with cluster-forming threshold $P_{unc.} < 0.001$). As shown by post hoc paired comparisons, this stage
514 difference was attributable to the increased correlation between dACC activation and RD values in S2
515 vs. S1, and S3 vs. S1 (Fig. 4D, left), but not by S2 vs. S3, in either positive or negative direction (Fig.
516 4D, right; see also Fig. 4B for ROI analyses). Together, these data provided consistent evidence
517 suggesting that the neural correlates of RD in S1 were spatially segregated from those in S2 and S3 in
518 a manner consistent with the DeGroot-RL prediction.

519
520 **dACC sensitivity to RD values was predictive of behavioral sensitivity to RD.** To relate the encoding
521 of RD values in the dACC to choice behavior, we tested whether, across subjects, the extent to which
522 dACC activity reflected RD values was predictive of the behavioral effects of RD on learning. We used
523 the individual value estimate of learning rate in S2 and S3 (i.e., β as in Fig. 1C) as a measure of how
524 strongly RD affected learning at these stages (zero effects on learning when $\beta = 0$). To capture the
525 overall individual neural sensitivity to RD, we averaged the dACC beta of RD from S2 and S3 in each
526 subject. We then plotted the individual behavioral estimate of β against the dACC beta in S2 and S3,
527 controlling for the respective baseline effects in S1 (Fig. 4E). The data showed that subjects with higher
528 learning rates in S2 and S3 than in S1 exhibited greater dACC sensitivity to RD values at observation
529 onsets in S2 and S3 than in S1 (Pearson's $r = 0.62$, $P = 0.001$). This between-subject association not
530 only held for dACC beta values averaged over S2 and S3, but was also highly significant and denoted
531 similar effect sizes when tested separately in these stages (S2: $r = 0.52$, $P = 0.007$; S3: $r = 0.53$, $P =$
532 0.006 ; S2 vs. S3: $\beta = -0.11 \pm 0.25$, $t_{24} = 0.43$, $P = 0.673$). Notably, this neural-behavioral association
533 was not a spurious effect arising from the double dipping of data⁴⁹, because the dACC beta of RD was
534 purely determined by the neural responses to the exogenously-given networks, independent of
535 participants choices, model specification, or data estimation.

536
537 **Ventromedial prefrontal cortex (VMPFC) signaled the value estimate of updated belief**
538 **expectation at the time of observation in S1, S2, and S3.** The above results thus raised the question
539 of how social observations from disparate neighbors were integrated in the brain to inform the
540 subsequent decision. Unlike previous learning experiments, where subjects typically make a choice
541 immediately after an observation, our experiment required participants to cache a sequence of social
542 information until they were asked to make a decision. Thus, a sensible strategy based on the DeGroot-
543 RL hypotheses would be to maintain an expectation about the unknown state and sequentially update
544 the expectation using either the unweighted (in S1) or RD-weighted (in S2 and S3) prediction error
545 signals each time an observation is witnessed.

546



547
 548 **Fig. 5 | Ventromedial prefrontal cortex (VMPFC) tracks the value estimate of updated belief expectation (E_{new}) in S1,**
 549 **S2, and S3. (A)** Statistical parametric map with respect to E_{new} estimates at observation onsets, computed by averaging the
 550 GLM5 coefficients for E_{new} across S1, S2, and S3 for each subject in the same way we looked for neural correlates of aPE
 551 estimates (MNI: $x, y, z = -9, 44, -10$; cluster-wise FWE-corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$;
 552 *Methods*). **(B)** Similar effect sizes in the parametric encoding of E_{new} estimates in the VMPFC across stages, as demonstrated
 553 by violin plots for the distribution of neural betas for E_{new} estimates. The neural betas were extracted for each separate stage
 554 from the significant VMPFC cluster as identified in Fig. 5A. **(C)** Increased functional connectivity between the seed region in
 555 the VMPFC and a cluster in the anterior cingulate cortex (ACC) at observation onsets in S2 and S3, relative to S1 (cluster-
 556 wise FWE-corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$). Similar to other GLM analyses in the study, this
 557 psychophysiological interaction (PPI) analysis compared the connectivity strength averaged over S2 and S3 against that in S1
 558 (*Methods*). The seed region was defined as a 6-mm sphere around the peak activation as identified in Fig. 5A. **(D)** No
 559 systematic difference in the effect sizes of functional coupling between S2 and S3 (paired t -test, $t_{24} = -1.20$, $P = 0.242$), as
 560 revealed by the PPI betas extracted from the significant cluster in the ACC as identified in Fig. 5C.

561
 562 This hypothesis immediately led to two neural predictions. First, signals reflecting the value estimate
 563 for updated expectation (E_{new} as in Fig. 1C) might be represented in brain regions previously
 564 implicated in tracking RL expectations, like the OFC or VMPFC^{13,14}. That is, in addition to the classic
 565 RL signals for belief expectations of the chosen option at choice time (as shown in Extended Data Fig.
 566 9A), we would also expect—at observation times—the neural representation of E_{new} estimate
 567 associated with the option previously selected by the observer that has been updated according to the
 568 observed action. Similar to aPE signals, we hypothesized that signals related to E_{new} estimates would
 569 be seen on an observation-by-observation basis and across three stages nonselectively. The second
 570 prediction was motivated by the DeGroot-RL hypothesis that, compared to S1, incorporating an aPE
 571 signal into belief expectation in S2 and S3 would involve additional modulatory inputs. Thus, regions
 572 representing E_{new} estimates might demonstrate increased functional connectivity in S2 and S3
 573 compared to S1, with regions related to tracking, representing, or implementing modulatory signals in
 574 service of learning.

575
 576 We tested the first prediction in a new GLM (GLM5, *Methods*), which included the value estimate of
 577 E_{new} as a parametric modulator at the time of observation onset of in the corresponding learning stage.
 578 To control for decision factors that might be related to learning or belief updating, we included the

579 following variables as regressors of no interest in the same regression model: aPE estimate associated
580 with the observed action, RD, and the order of observation display. Similar to the above analyses for
581 aPE estimates, we averaged each individual's regression coefficients with respect to E_{new} estimates
582 across three stages and then took them into second-level analyses. We found a strong positive
583 correlation between E_{new} estimates and activity in a number of brain regions including the VMPFC
584 (Fig. 5A; cluster-wise FWE corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$;
585 Supplementary Table 5). The observed VMPFC responses to E_{new} estimates were robust (cluster-wise
586 FWE-corrected $P < 0.05$, with cluster-forming threshold $P_{unc.} < 0.001$) to the inclusion of additional
587 decision-related variables, such as the product of the RD value and aPE estimate, color selected by the
588 observee, visual properties related to the network display (visual centrality and distance), and the global
589 consensus level at the beginning of the learning stage (GLM6, *Methods*). As predicted, we observed
590 stable neural representation of E_{new} estimates across S1, S2, and S3, such that within-subject
591 comparisons identified no significant difference in neural responses to E_{new} estimates across stages
592 either within the VMPFC cluster (Fig. 5B; one-way repeated-measures ANOVA, $F(2,48) = 2.33$, $P =$
593 0.108) or at the whole-brain level (cluster-wise FWE-corrected $P < 0.05$, with cluster-forming threshold
594 $P_{unc.} < 0.001$).

595
596 To test the second prediction, we performed an exploratory psychophysiological interaction (PPI)
597 analysis to look for brain regions that showed differential functional coupling with the VMPFC when
598 RD was vs. was not needed for scaling aPE signals. Seeded in the VMPFC (6 mm around the peak
599 activation as identified in Fig. 5A), the PPI analysis compared the connectivity strength averaged over
600 S2 and S3 against that in S1 (*Methods*). This showed increased coupling of the VMPFC with several
601 regions including a cluster in the ACC (Fig. 5C; cluster-wise FWE-corrected $P < 0.05$, with cluster-
602 forming threshold $P_{unc.} < 0.001$), with no significant difference in the coupling effect sizes between S2
603 and S3 (Fig. 5D; paired difference for S2 – S3 = -0.03 ± 0.02 , $t_{24} = -1.20$, $P = 0.242$). The cluster
604 identified by the PPI analysis partially overlapped with the area in the dACC signaling RD values in S2
605 and S3 (Extended Data Fig. 13), yet its peak activation was located in the more ventral and rostral
606 portion of the ACC (MNI: $x, y, z = 3, 20, 23$; Fig. 5C). On the one hand, the overlapping dACC
607 activation points to a possibility that the dACC might be involved in both representing the degree-
608 related information and using this information for modulating the VMPFC representation, when
609 network features are relevant for learning. On the other hand, as the ACC is a richly intra-connected
610 system with projections to a broad set of regions, including the VMPFC⁵⁰, it is also possible that the
611 dACC cluster identified for encoding RD played an indirect role in influencing the VMPFC, through
612 its effects on other parts of the ACC, such as its ventral and rostral portions, that have been previously
613 implicated in monitoring and integrating learning signals^{19,42}.

614

615 Discussion

616 Information flowing in a large-scale, interconnected society is often entangled, conflated, and
617 sometimes superfluous^{1,26}. This poses a computational challenge for social learning, during which
618 agents need to reconcile disparate sources of signals based on their informativeness⁵¹. Prior research on
619 individual learning in non-social contexts has shown that humans can accurately estimate how relevant
620 a learning signal is in predicting future and use this estimate to adjust RL learning rates^{42,52,53}. On social
621 networks, however, optimally evaluating the predictive value of each observation is cognitively
622 demanding, sometimes even prohibitive. Indeed, failure to effectively aggregate information from

623 connected peers has long been hypothesized to underlie herding, social influence biases, misinformation
624 propagation, and other forms of collective maladaptation^{9,54}.

625

626 Combining fMRI, formal theories of RL, and social network analysis, we explored the possibility that
627 to balance computational costs, the brain approximates the relative informativeness of a social signal
628 based on the structural properties of the network that routes information transmission. To evaluate this
629 possibility, we grounded the DeGroot learning heuristics, a classic theory for naïve social learning
630 imported from social network analysis^{7,30,31}, into the temporal difference form of RL, widely implicated
631 in the neurobiology of learning and decision-making^{21,42,52}. Using a real-time, distributed learning task
632 on networks with varying topological structures, this study provided behavioral and neural evidence
633 that learning in complex, interconnected environments can be realized by means similar to the well-
634 established RL algorithm. Importantly, the RL learning rate fluctuated according to a signal related to
635 network degree centrality, indexed by dACC activity at the time of witnessing others' actions, but only
636 insofar as the social observations varied in their informativeness.

637

638 The observed dACC response to the degree centrality of the observee relative to that of the observer
639 (i.e., RD) in S2 and S3 is consistent with past evidence showing a key role of this region in facilitating
640 behavioral flexibility and adjusting learning rates for adapting to the external world⁴¹⁻⁴⁵. Our data extend
641 these findings by demonstrating that the regulatory process may also incorporate the topological
642 properties of social connections that underlie learning. In our case, the dACC encoding was seen on an
643 observation-by-observation basis, reflected the degree centrality of the observee and observer
644 simultaneously and independently, existed above and beyond the prediction error and other decision-
645 related variables, and its across-subject variations were predictive of individual differences in the degree
646 modulation effect on behavior.

647

648 Importantly, our data also emphasized the absence of a dACC response to the same RD signal in S1
649 when the network structure was irrelevant to learning. This finding argues against the possibility that
650 the dACC engagement identified in S2 and S3 was due to some low-level visual processing of the
651 network displays, or due to other more general functions of the dACC—such as detecting errors⁵⁵ or
652 monitoring social conflicts⁵⁶—that would be involved across all learning stages nonselectively.
653 Alternatively, the stage-dependent dACC encoding is consistent with a broader proposal of this region,
654 suggested by past neurophysiological and neuroimaging evidence, as representing task-relevant (but
655 not irrelevant) information that supports behavioral changes and guides appropriate action selection⁴¹.
656 Our data thus suggest the involvement of a high-level, controlled process in evaluating the source of
657 social information in service of learning, and argue against a model of blind, automatic discrimination
658 among social contacts in explaining social information aggregation. More broadly, the observed
659 between-stage differences echo past studies that used data from social media and highlighted the
660 importance of separating information propagation stages, such as those related to the initial transmission
661 and retransmission, in developing mechanistic understandings for rumor dissemination and
662 amplification^{26,57}.

663

664 While our findings highlighted a role of the dACC specific to S2 and S3, we also observed that, in S1,
665 activity in the PCC, precuneus, and visual cortex correlated with measures of degree centralities. There
666 are several possibilities for how network-related activity in S1 would contribute to learning. One
667 possibility is that the S1 activation is associated with the recognition or representation of network
668 features, which facilitates the flexible usage of those features in the latter stages. Indeed, the loci of the

669 S1 activation were similar to those implicated in representing the internal perception of centralities or
670 other characters of real-life social networks, when subjects were required to view pictures or videos of
671 their acquaintances^{29,47,48}. Our data are thus consistent with a possibility that, whereas the degree-
672 modulation effect is stage dependent, some network-related information may be automatically
673 registered in the brain to prepare for the future usage. Alternatively, it is also possible that, S1 activation
674 may reflect, rather than the perception of interpersonal connections, some low-level processing of the
675 network stimuli (e.g., visual processing), which typically shows a more pronounced activation when
676 the stimulus is novel than when the stimulus has been recently processed⁵⁸. To clarify these possibilities
677 and explore how network structure is internally perceived, represented, and transformed into the
678 modulation signal in service of learning, future research is needed to combine the current approach with
679 sociometric methods used for studying real-world connections⁵⁹, to investigate how the brain learns
680 from actual peers without resorting to networks that are artificially structured and displayed.

681

682 Degree centrality has long been hypothesized to have a close relationship with social influence in small
683 group interaction and communication⁶⁰. Our finding that the brain modulated learning according to
684 degree-related signals is consistent with two broad accounts previously proposed for how network
685 centrality affects learning. The first has its basis in human and non-human studies that emphasize the
686 role of the structural position in social behavior, suggesting that structurally comparable individuals are
687 facing similar interacting environments, therefore exhibiting similar behavior toward one another^{25,28}.
688 In the context of learning, this account proposes that the opportunity to obtain new information through
689 interpersonal interactions may be constrained by one's location on the interaction network, and can be
690 quantified by location features, such as the degree centrality. A second but not mutually exclusive
691 possibility has its basis in the dynamic nature of network topology: Knowledgeable or successful
692 individuals tend to become highly connected, thus degree centrality may serve to signal an individual's
693 capability or social status to other individuals^{28,61}. Under this possibility, social animals may have
694 evolved to preferentially follow the more "connected" or "prestigious" conspecifics, even in controlled
695 experiments where the network structure is fixed and locations are randomized.

696

697 Compatible with these diverse lines of proposals, our data additionally highlighted a dual effect of
698 centrality on learning: Higher degree centrality not only amplified one's social influence, but also
699 reduced one's susceptibility to others' influence. This finding is consistent with the behavioral evidence
700 from popular social media, demonstrating that more influential individuals are usually less susceptible
701 to peers' influence, compared to their less influential counterparts³⁴. Our results thus point to an exciting
702 possibility that, while social influence and susceptibility to social influence are often considered as
703 distinct personal attributes⁵⁶, they may be jointly affected by an internal learning system, which
704 approximates the predictive value of others' information relative to one's own, in order to cope with the
705 complexity of social environments.

706

707 It is worth noting that, owing to the fundamental role of degree centrality in network analyses and its
708 close relationship with a range of network characteristics, we cannot rule out the possibility that
709 alternative network features may contribute to learning. For example, in addition to the degree centrality,
710 which parameterizes the immediate effect of social influence, learning may be affected by measures
711 such as the eigenvector centrality, closeness, clustering coefficient, betweenness, or constraint
712 coefficient, which have been used to examine information propagation from the perspective of long-
713 term, sequential, circular, globally or locally mediating effect, respectively². Our focus on degree
714 centrality reflects the assumption that the brain may be more sensitive to simple, straightforward

715 geometric properties, especially for complex decisions. Indeed, across analyses, there was no evidence
716 that alternative metrics outperformed RD in explaining either the behavioral data or the dACC
717 responses. Future investigation is needed to more firmly isolate and compare the potential influences
718 of various network features at the behavioral and neural levels.

719

720 Previous research on the neurocomputational processes of social learning has typically focused on
721 highly simplified interpersonal settings, leaving open whether and how putative RL mechanisms
722 identified in simplistic setups can support behavior in more complex, ecologically-relevant
723 environments. Here, we showed that key features of learning in interconnected contexts were consistent
724 with an error-driven process, similar to those seen in nonnetworked situations. As the network structures
725 examined in the present study were but a sample of immense possibilities of real-world social networks,
726 these results raise questions regarding the scalability and generality of the proposed model. First, the
727 experiment focused on relatively small, 7-node networks, and did not directly speak to larger, more
728 naturalistic settings. Nonetheless, we speculate that, by relying merely on local information, the
729 proposed model may be particularly suitable for scaling up, as individuals in large social groups
730 typically only have access to the local knowledge but not the global information such as the structure
731 of the entire network. Alternative RL algorithms for learning in large-scale networked systems have
732 been developed in control engineering⁶². These algorithms usually aim at optimizing some global
733 network performances (e.g., total reward) and their cognitive and neurobiological feasibility is yet to
734 be evaluated.

735

736 Second, the DeGroot-RL model explained choice behavior and task-related neural activity,
737 demonstrating no systematic differences in its explanatory power across networks (Extended Data Fig.
738 14). Yet, it remains possible that the brain may follow other learning algorithms when facing a different
739 set of networks—for example, deploying Bayesian strategies when making decisions in a line, one of
740 the simplest forms of directed network³⁷. Hybrid learning is also possible according to a recent
741 behavioral study suggesting a mixture of Bayesian and DeGroot learning in a relatively more educated
742 (but not less educated) sample³⁶. The current study constitutes an initial step toward a neural
743 mechanistic understanding of learning on social networks. Future studies are needed to address whether,
744 and under what circumstances, our findings can be extended to study the potential involvement of
745 multiple learning systems, arbitration between those systems, and individual differences in related
746 processes. Learning on networks offers an excellent opportunity for probing the influence of social
747 structure on the internal tradeoff between computational complexity and learning effectiveness^{63,64}.

748

749 Social networks have been widely hypothesized to play a key role in many large-scale social phenomena,
750 including vaccine hesitancy, voting behavior, and fake news proliferation, yet the exact mechanisms by
751 which interpersonal connections contribute to these phenomena remain unclear. The current study sheds
752 light on this topic from a neurocognitive perspective, by elucidating how individuals actually
753 experience and interact with a networked environment. Our data provide neural evidence for a bounded
754 rational, network-related filtering of social information, which may result in the spread of
755 misinformation and biased consensus among connected peers. More broadly, this work demonstrates
756 the possibility of developing computationally-tractable and neurobiologically-plausible tools and
757 methods for investigating the complex interplay between social behavior and social embedding in the
758 brain, which may have the potential to translate upward for tackling phenomena in wider society.

759

760 References

- 761 1. Centola, D. *How Behavior Spreads*. (Princeton University Press, 2020).
- 762 2. Wasserman, S. & Faust, K. *Social Network Analysis*. (Cambridge University Press, 1994).
- 763 3. Momennejad, I. Collective minds: social network topology shapes collective cognition. *Philos*
764 *Trans R Soc Lond B Biol Sci* **377**, 20200315 (2022).
- 765 4. Buyalskaya, A., Gallo, M. & Camerer, C. F. The golden age of social science. *Proc Natl Acad*
766 *Sci USA* **118**, e2002923118 (2021).
- 767 5. Falk, E. B. & Bassett, D. S. Brain and social networks: fundamental building blocks of human
768 experience. *Trends in Cognitive Sciences* **21**, 674–690 (2017).
- 769 6. Stewart, A. J. *et al.* Information gerrymandering and undemocratic decisions. *Nature* **573**, 1–9
770 (2019).
- 771 7. Jackson, M. O. *Social and Economic Networks*. (Princeton University Press, 2010).
- 772 8. Christakis, N. A. & Fowler, J. H. *Connected*. (Little, Brown Spark, 2009).
- 773 9. Raafat, R. M., Chater, N. & Frith, C. Herding in humans. *Trends in Cognitive Sciences* **13**,
774 420–428 (2009).
- 775 10. Pennycook, G. & Rand, D. G. The Psychology of Fake News. *Trends in Cognitive Sciences*
776 **25**, 388–402 (2021).
- 777 11. Sutton, R. S. & Barto, A. G. *Reinforcement Learning*. (MIT Press, 1998).
- 778 12. Schultz, W., Dayan, P. & Montague, P. R. A Neural Substrate of Prediction and Reward.
779 *Science* **275**, 1593–1599 (1997).
- 780 13. Niv, Y. Reinforcement learning in the brain. *Journal of Mathematical Psychology* **53**, 139–154
781 (2009).
- 782 14. Glimcher, P. W. Understanding dopamine and reinforcement learning: The dopamine reward
783 prediction error hypothesis. *Proc Natl Acad Sci USA* **108**, 15647–15654 (2011).
- 784 15. King-Casas, B. *et al.* Getting to Know You: Reputation and Trust in a Two-Person Economic
785 Exchange. *Science* **308**, 78–83 (2005).
- 786 16. Behrens, T. E. J., Hunt, L. T., Woolrich, M. W. & Rushworth, M. F. S. Associative learning of
787 social value. *Nature* **456**, 245–249 (2008).
- 788 17. Hampton, A. N., Bossaerts, P. & O’Doherty, J. P. Neural correlates of mentalizing-related
789 computations during strategic interactions in humans. in **105**, 6741–6746 (2008).
- 790 18. Zhu, L., Mathewson, K. E. & Hsu, M. Dissociable neural representations of reinforcement and
791 belief prediction errors underlie strategic learning. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1419–
792 1424 (2012).
- 793 19. Suzuki, S., Adachi, R., Dunne, S., Bossaerts, P. & O’Doherty, J. P. Neural mechanisms
794 underlying human consensus decision-making. *Neuron* **86**, 591–602 (2015).
- 795 20. Zhang, L. & Gläscher, J. A brain network supporting social influences in human decision-
796 making. *Science Advances* **6**, eabb4159 (2020).
- 797 21. Burke, C. J., Tobler, P. N., Baddeley, M. & Schultz, W. Neural mechanisms of observational
798 learning. *Proc Natl Acad Sci USA* **107**, 14431–14436 (2010).
- 799 22. Suzuki, S. *et al.* Learning to simulate others' decisions. *Neuron* **74**, 1125–1137 (2012).
- 800 23. Campbell-Meiklejohn, D., Simonsen, A., Frith, C. D. & Daw, N. D. Independent neural
801 computation of value from other people's confidence. *Journal of Neuroscience* **37**, 673–684
802 (2017).
- 803 24. Dunne, S. & O’Doherty, J. P. Insights from the application of computational neuroimaging to
804 social neuroscience. *Current Opinion in Neurobiology* **23**, 387–392 (2013).
- 805 25. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: homophily in social
806 networks. *Annual Review of Sociology* **27**, 415–444 (2001).
- 807 26. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. The
808 echo chamber effect on social media. *Proc. Natl. Acad. Sci. U.S.A.* **118**, (2021).
- 809 27. DeMarzo, P. M., Vayanos, D. & Zwiebel, J. Persuasion bias, social influence, and
810 unidimensional opinions. *The Quarterly Journal of Economics* **118**, 909–968 (2003).
- 811 28. Kulahci, I. G. & Quinn, J. L. Dynamic relationships between information transmission and
812 social connections. *Trends in Ecology & Evolution* 1–10 (2019).

- 813 29. Parkinson, C., Kleinbaum, A. M. & Wheatley, T. Spontaneous neural encoding of social
814 network position. *Nat. hum. behav.* **1**, 1–7 (2017).
- 815 30. Mobius, M. & Rosenblat, T. Social learning in economics. *Annu. Rev. Econ.* **6**, 827–847
816 (2014).
- 817 31. DeGroot, M. H. Reaching a consensus. *Journal of the American Statistical Association* **69**,
818 118–121 (1974).
- 819 32. Golub, B. & Jackson, M. O. Naïve learning in social networks and the wisdom of crowds.
820 *American Economic Journal: Microeconomics* **2**, 112–149 (2010).
- 821 33. Molavi, P., Tahbaz-Salehi, A. & Jadbabaie, A. A theory of non-Bayesian social learning.
822 *Econometrica* **86**, 445–490 (2018).
- 823 34. Aral, S. & Walker, D. Identifying influential and susceptible members of social networks.
824 *Science* **337**, 337–341 (2012).
- 825 35. Grimm, V. & Mengel, F. Experiments on belief formation in networks. *Journal of the*
826 *European Economic Association* **9**, 41 (2018).
- 827 36. Chandrasekhar, A. G., Larreguy, H. & Xandri, J. P. Testing models of social learning on
828 networks: evidence from two experiments. *Econometrica* **88**, 1–32 (2020).
- 829 37. Anderson, L. R. & Holt, C. A. Information cascades in the laboratory. *The American*
830 *Economic Review* **87**, 847–862 (1997).
- 831 38. Toelch, U. & Dolan, R. J. Informational and normative influences in conformity from a
832 neurocomputational perspective. *Trends in Cognitive Sciences* **19**, 579–589 (2015).
- 833 39. Ramsey, R., Kaplan, D. M. & Cross, E. S. Watch and learn: the cognitive neuroscience of
834 learning from others' actions. *Trends in Neurosciences* **44**, 478–491 (2021).
- 835 40. Kolling, N., Behrens, T., Wittmann, M. K. & Rushworth, M. Multiple signals in anterior
836 cingulate cortex. *Current Opinion in Neurobiology* **37**, 36–43 (2016).
- 837 41. Heilbronner, S. R. & Hayden, B. Y. Dorsal anterior cingulate cortex: a bottom-up view.
838 *Annual Review of Neuroscience* **39**, 149–170 (2016).
- 839 42. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value
840 of information in an uncertain world. *Nat Neurosci* **10**, 1214–1221 (2007).
- 841 43. Jocham, G., Neumann, J., Klein, T. A., Danielmeier, C. & Ullsperger, M. Adaptive coding of
842 action values in the human rostral cingulate zone. *J. Neurosci.* **29**, 7489–7496 (2009).
- 843 44. O'Reilly, J. X. *et al.* Dissociable effects of surprise and model update in parietal and anterior
844 cingulate cortex. *Proc Natl Acad Sci USA* **110**, E3660–E3669 (2013).
- 845 45. Sheth, S. A. *et al.* Human dorsal anterior cingulate cortex neurons mediate ongoing
846 behavioural adaptation. *Nature* **488**, 218–221 (2012).
- 847 46. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale
848 automated synthesis of human functional neuroimaging data. *Nature Methods* **8**, 665–670
849 (2011).
- 850 47. Zerubavel, N., Bearman, P. S., Weber, J. & Ochsner, K. N. Neural mechanisms tracking
851 popularity in real-world social networks. *Proc Natl Acad Sci USA* **112**, 15072–15077 (2015).
- 852 48. Morelli, S. A., Leong, Y. C., Carlson, R. W., Kullar, M. & Zaki, J. Neural detection of socially
853 valued community members. *Proc Natl Acad Sci USA* **15**, 201712811–6 (2018).
- 854 49. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F. & Baker, C. I. Circular analysis in
855 systems neuroscience: the dangers of double dipping. *Nat Neurosci* **12**, 535–540 (2009).
- 856 50. Apps, M. A. J., Rushworth, M. F. S. & Chang, S. W. C. The anterior cingulate gyrus and
857 social cognition: tracking the motivation of others. *Neuron* **90**, 692–707 (2016).
- 858 51. FeldmanHall, O. & Nassar, M. R. The computational challenge of social learning. *Trends in*
859 *Cognitive Sciences* 1–13 (2021).
- 860 52. Dayan, P., Kakade, S. & Montague, P. R. Learning and selective attention. *Nat Neurosci* **3**,
861 1218–1223 (2000).
- 862 53. Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V. & Niv, Y. Dynamic Interaction
863 between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*
864 **93**, 451–463 (2017).
- 865 54. Lazer, D. M. J. *et al.* The science of fake news. *Science* **359**, 1094–1096 (2018).
- 866 55. Carter, C. S. & van Veen, V. Anterior cingulate cortex and conflict detection: an update of
867 theory and data. *Cognitive, Affective, & Behavioral Neuroscience* **7**, 367–379 (2007).

- 868 56. Cascio, C. N., Scholz, C. & Falk, E. B. Social influence and the brain: persuasion,
869 susceptibility to influence and retransmission. *Current Opinion in Behavioral Sciences* **3**, 51–
870 57 (2015).
- 871 57. Sutton, J. *et al.* A cross-hazard analysis of terse message retransmission on Twitter. *Proc Natl*
872 *Acad Sci USA* **112**, 14793–14798 (2015).
- 873 58. Grill-Spector, K., Henson, R. & Martin, A. Repetition and the brain: neural models of
874 stimulus-specific effects. *Trends in Cognitive Sciences* **10**, 14–23 (2006).
- 875 59. Galesic, M. *et al.* Human social sensing is an untapped resource for computational social
876 science. *Nature* **595**, 214–222 (2021).
- 877 60. Friedkin, N. E. *A Structural Theory of Social Influence*. (Cambridge University Press, 1998).
- 878 61. Almaatouq, A., Noriega-Campero, A., Alotaibi, A., Krafft, P. M. & Pentland, A. Adaptive
879 social networks promote the wisdom of crowds. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11379–
880 11386 (2020).
- 881 62. Qu, G., Wierman, A. & Li, N. Scalable reinforcement learning of localized policies for multi-
882 agent networked systems. *arXiv.org math.OC*, arXiv:1912.02906 (2019).
- 883 63. Charpentier, C. J., Iigaya, K. & O’Doherty, J. P. A neuro-computational account of arbitration
884 between choice imitation and goal emulation during human observational learning. *Neuron*
885 **106**, 687–699.e7 (2020).
- 886 64. Bossaerts, P. & Murawski, C. Computational complexity and human decision-making. *Trends*
887 *in Cognitive Sciences* **21**, 917–929 (2017).
- 888

889 **Data availability**

890 Data underlying the findings of this study will be available on the Open Science Framework upon
891 acceptance.

892

893 **Code availability**

894 Code supporting the findings of this study will be available on the Open Science Framework upon
895 acceptance.

896

897 **Acknowledgements**

898 We thank Yunlu Yin and Yan Wang for assistance with intranet setup and data collection. We also thank
899 the National Center for Protein Sciences and the high-performance computing platform at the Center
900 for Life Sciences at Peking University for facilitating data acquisition and computation. This work is
901 supported by NSFC (32071095 to L.Z.) and Center for Life Sciences at Peking University.

902

903 **Author contributions**

904 Y.J. and L.Z. designed the study. Y.J. conducted the experiments. All authors analyzed the data and
905 wrote the manuscript.

906

907 **Competing interests**

908 The authors declare no competing interests.