



# Aspiration dynamics generate robust predictions in heterogeneous populations

Lei Zhou <sup>1,2</sup>, Bin Wu <sup>3</sup>, Jinming Du <sup>4,5</sup> & Long Wang <sup>1</sup>✉

Update rules, which describe how individuals adjust their behavior over time, affect the outcome of social interactions. Theoretical studies have shown that evolutionary outcomes are sensitive to model details when update rules are imitation-based but are robust when update rules are self-evaluation based. However, studies of self-evaluation based rules have focused on homogeneous population structures where each individual has the same number of neighbors. Here, we consider heterogeneous population structures represented by weighted networks. Under weak selection, we analytically derive the condition for strategy success, which coincides with the classical condition of risk-dominance. This condition holds for all weighted networks and distributions of aspiration levels, and for individualized ways of self-evaluation. Our findings recover previous results as special cases and demonstrate the universality of the robustness property under self-evaluation based rules. Our work thus sheds light on the intrinsic difference between evolutionary dynamics under self-evaluation based and imitation-based update rules.

<sup>1</sup>Center for Systems and Control, College of Engineering, Peking University, Beijing, China. <sup>2</sup>Center for Research on Intelligent System and Engineering, Institute of Automation, Chinese Academy of Sciences, Beijing, China. <sup>3</sup>School of Sciences, Beijing University of Posts and Telecommunications, Beijing, China. <sup>4</sup>Institute of Industrial and Systems Engineering, College of Information Science and Engineering, Northeastern University, Shenyang, China. <sup>5</sup>Key Laboratory of Data Analytics and Optimization for Smart Industry (Northeastern University), Ministry of Education, Shenyang, China.  
✉email: [longwang@pku.edu.cn](mailto:longwang@pku.edu.cn)

Social behaviors such as cooperation are deeply rooted in the daily interactions across all levels of organisms. In particular, humans exhibit remarkable capabilities of cooperating, coordinating, and dividing tasks with other humans. Such social behaviors are indispensable to the survival and development of human societies. Understanding the evolution of social behavior is thus of great importance<sup>1,2</sup>. After decades of investigations, evolutionary game theory has been demonstrated as a powerful tool to study the evolution of social behavior<sup>3,4</sup>.

Indeed, past decades have seen intensive investigations of evolutionary games in structured populations<sup>5–9</sup>. One of the most important questions is how population structure alters evolutionary outcomes. It is shown that the answer to this question strongly depends on update rules<sup>7,10–12</sup>. Update rules are explicit behavioral rules of individuals, which specify what kind of information they use and how they process such information to determine future behaviors<sup>7</sup>. In evolutionary games, the information required by update rules usually includes individuals' strategies and payoffs. As the input for decision making, the information used is likely to affect individuals' behavioral updating, resulting in changes at the population level. One kind of information that receives particular attention is social peers' payoff information. Recent human behavioral experiments suggest that whether individuals use social peers' payoff information to update behavior may be crucial to the outcome of social interactions, for example, the level of cooperation in groups<sup>13,14</sup> or on network-structured populations<sup>15–18</sup>.

Based on the relevance of social peers' payoffs, update rules in theoretical models can be classified into two classes: imitation-based (relevant) and self-evaluation based (irrelevant). Under imitation-based rules, individuals update strategies by copying more successful peers. When using self-evaluation based rules, individuals self-assess performance of strategies and then switch to strategy alternatives<sup>7</sup>. Self-evaluation can be based on aspirations: individuals compare payoffs with their aspirations and then switch based on the shortfall of payoffs<sup>19–23</sup>. Update rules of these two classes are both common in practice and they are tailored for different environment. For example, if individuals are not confident to make decisions or uncertain about the consequences, imitating the more successful provides valuable shortcuts for decision-makers. Self-evaluation, instead, is efficient and superior when social information is unavailable, regarded as unreliable, or costs individuals too much to gather and process.

It is well known that imitation-based update rules lead to evolutionary outcomes sensitive to model details, such as population structures<sup>11,12,24</sup>, the way of imitating<sup>10,25</sup>, and heterogeneity of decision-making rules<sup>26,27</sup>. Such sensitivity makes it difficult for researchers to generalize predictions across different population structures (e.g., from regular to non-regular networks) or different imitation-based rules (e.g., from death-birth<sup>28</sup> to pairwise comparison rules<sup>29,30</sup>). On the other hand, self-evaluation based rules are shown to generate robust evolutionary outcomes. For instance, the condition for strategy success on unweighted regular networks is found to be the same as that in well-mixed populations<sup>31</sup>, and such robustness is not affected by different distributions of aspiration values<sup>32</sup> or heterogeneous ways of self-evaluation<sup>33</sup>. Despite the findings on unweighted regular networks<sup>31–33</sup>, it remains unclear whether the robustness property under self-evaluation based rules applies to heterogeneous population structures, where the number of neighbors varies from one individual to another and each individual interacts with their neighbors under different rates.

To fill this gap, we study evolutionary games under aspiration-based self-evaluation rules (for short, aspiration dynamics) on heterogeneous population structures represented by weighted networks. Under the limit of weak selection and symmetric

aspirations, we analytically derive a condition for one strategy to prevail over the other, which is found to coincide with the classical condition of risk-dominance. This condition holds for any weighted network, any distribution of aspiration levels, and for any individualized ways of self-evaluation. If aspirations are differentiated by strategies, we find that the condition of risk-dominance is altered and cooperation can evolve in the Prisoner's Dilemma game if individuals aspire more when they defect. The intuitive interpretation of our results is as follows: one strategy prevails over the other if the strategy on average brings more satisfaction to individuals than the other does. Our work thus (i) highlights that switching off from social peers' payoff information while updating strategies has a nontrivial impact on the evolutionary outcomes, and (ii) demonstrates that the robustness property of aspiration-based self-evaluation rules is universal in heterogeneous populations.

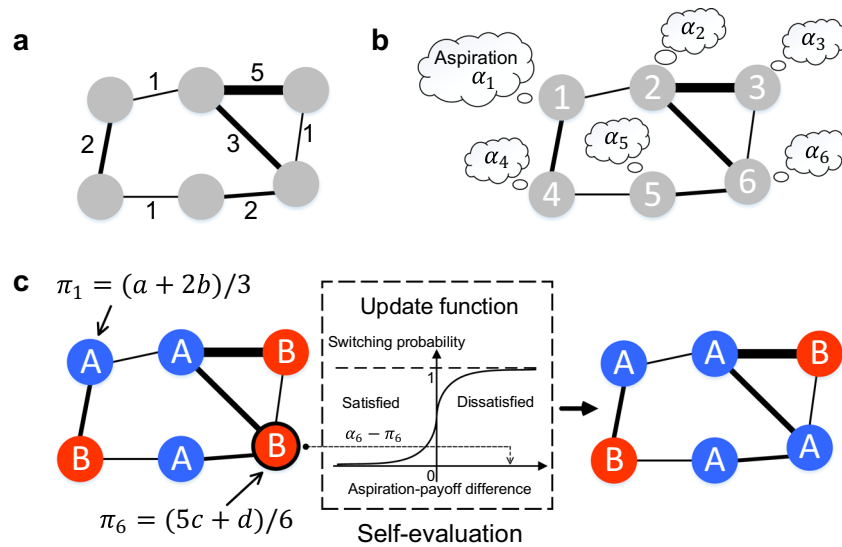
## Results

**Population structure and games.** We consider a population with fixed size  $N$  ( $N \geq 2$ ). The population structure is depicted by a static weighted graph (or network) with edge weights  $w_{ij} \geq 0$ , where vertices represent individuals, edges indicate who interacts with whom, and weights describe the number of interactions per unit time. Self-interactions are excluded. Individuals collect edge-weighted average payoffs by playing games with their nearest neighbors<sup>24</sup>. The total number of interactions each individual  $i$  engages in is  $d_i = \sum_{j=1}^N w_{ij}$  ( $i = 1, 2, \dots, N$ ). We assume  $d_i > 0$  for all  $i$ , which means that each individual has at least one neighbor to interact with. Visually, the graph should have neither isolated vertices nor self-loops, which are natural assumptions when studying evolutionary games on graphs. In each game, individuals play either strategy  $A$  or strategy  $B$ . The payoff matrix of the game is given by

$$\begin{array}{cc} & \begin{array}{c} A \\ B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{array}, \quad (1)$$

where both players get payoff  $a$  if they play strategy  $A$  ( $A$ -player) and get  $d$  if they play strategy  $B$  ( $B$ -player); if an  $A$ -player encounters a  $B$ -player, the former obtains payoff  $b$  and the latter  $c$ . For each individual  $i$ , we denote  $\pi_{i,X}$  as its payoff when it uses strategy  $X$  ( $X = A, B$ ).

**Aspiration dynamics.** At each time step, an individual is randomly selected and given the opportunity to revise its strategy. We assume individuals follow self-evaluation-based rules, under which they evaluate their strategies by comparing payoffs garnered from the games with their aspirations. Aspirations are either personalized<sup>32</sup>, which means each individual  $l$  has its own aspiration  $\alpha_l$  ( $l = 1, 2, \dots, N$ ), or contingent on strategies, which means that individuals using strategy  $A$  have an aspiration  $\alpha_A$  and those using  $B$  have  $\alpha_B$ . For simplicity, we consider fixed aspirations, meaning that there is no adaptation of aspirations due to learning. If such aspiration-driven update rules are deterministic, the aspiration level serves as a sharp boundary between satisfaction and disappointment<sup>34</sup>: if an individual's payoff exceeds its aspiration, the outcome is deemed satisfactory and it will repeat its strategy; if the payoff is otherwise lower than the aspiration, it feels disappointed and will switch to the other strategy. In real-life situations, strategy updating involves mistakes and admits bounded rationality, which is better captured by probabilistic (stochastic) strategy switchings. The probability can be determined by the level of satisfaction, i.e., the difference between the payoff and the aspiration. In our model, to be consistent with previous work<sup>32,33,35</sup>, we use update functions  $g: \mathbb{R} \rightarrow [0, 1]$  to



**Fig. 1** Aspiration dynamics on weighted graphs. **a** An undirected weighted graph with edge weight  $w_{ij} \geq 0$ . **b** Individuals occupy vertices of the graph and each individual  $i$  has an imaginary payoff value  $\alpha_i$  they aspire, called aspiration level. **c** For aspiration dynamics, at each time step, an individual is randomly selected (here, the sixth individual, marked by the black circle). It garners an edge-weighted average payoff ( $\pi_6$ ) by playing games with its nearest neighbors<sup>24</sup>. Then it self-evaluates the performance of the strategy in use by calculating the aspiration-payoff difference ( $\alpha_6 - \pi_6$ ), which is later used by the update function  $g: \mathbb{R} \rightarrow [0, 1]$  to determine its switching probability. If the payoff exceeds the aspiration, it feels satisfied and is more likely to keep its current strategy; otherwise, it is prone to switch. As illustrated,  $\alpha_6 - \pi_6 > 0$  (i.e.,  $\pi_6 < \alpha_6$ ) and the corresponding individual switches from strategy  $B$  to  $A$ .

map the aspiration-payoff difference into the switching probability (see Methods). We allow individuals to have their own update functions  $g_i$  since they may behave differently even for the same aspiration-payoff difference (i.e., level of dissatisfaction). Albeit this flexibility, all update functions should ensure that individuals have decreasing tendency to switch a strategy if it brings more satisfaction (see Methods) (Fig. 1).

Here, we employ stochastic self-evaluation-based rules. Under stochastic rules, the strictness of the strategy evaluation, namely, how much the payoff-aspiration difference affects individuals' decision-making, is controlled by the selection intensity  $\beta \geq 0$ <sup>36,37</sup>. Since each individual on the network uses either strategy  $A$  or  $B$ , the number of all possible states of the system is  $M = 2^N$ . Meanwhile, the transition probabilities between all the states can be described by an  $M \times M$  matrix  $\mathbf{P}_\beta$ . Similar to the mutation-selection process<sup>11,38</sup>, the resulting aspiration dynamics admit a unique stationary distribution  $\mathbf{u}_\beta$  (a column vector with  $M$  elements), which is the unique solution to the equation  $\mathbf{u}_\beta^T \mathbf{P}_\beta = \mathbf{u}_\beta^T$  (the superscript  $T$  represents vector/matrix transpose). In this distribution, we compare the average abundance (i.e., frequency) of strategy  $A$ ,  $\langle x_A \rangle = \mathbf{u}_\beta^T \mathbf{x}$ , with that of  $B$ ,  $\langle x_B \rangle = \mathbf{u}_\beta^T (\mathbf{1} - \mathbf{x})$ , where  $\mathbf{x}$  is the frequency of strategy  $A$  in each of the  $M$  states. If  $\langle x_A \rangle > \langle x_B \rangle$ , strategy  $A$  prevails over  $B$ . Otherwise,  $B$  prevails over  $A$ . We derive the condition for strategy success, i.e., the condition which leads to  $\langle x_A \rangle > \langle x_B \rangle$  or  $\langle x_B \rangle > \langle x_A \rangle$ . To make progress, we consider weak selection (i.e.,  $0 < \beta \ll 1$ )<sup>24,36,37,39-41</sup>, under which individuals switch strategies with a nearly constant probability. Weak selection may arise for the following reasons: (i) individuals have no obvious preference over different strategies<sup>42,43</sup>, (ii) individuals are uncertain about the payoffs, aspirations, or aspiration-payoff differences due to noise or stochastic interactions<sup>44</sup>.

**General condition for strategy success.** Given our assumptions, we calculate the average frequency difference between strategy  $A$  and  $B$ ,  $\langle x_A - x_B \rangle$ , in the stationary regime. If  $\langle x_A - x_B \rangle > 0$ ,

strategy  $A$  prevails over  $B$ ; otherwise, strategy  $B$  prevails over  $A$ . At the neutral drift  $\beta = 0$ , strategy  $A$  and strategy  $B$  are of equal abundance<sup>32</sup>. Under weak selection  $0 < \beta \ll 1$ , we use perturbation theory and get that  $\langle x_A - x_B \rangle > 0$  if

$$\mathbf{u}_0^T \mathbf{P}'_0 \mathbf{c} > 0, \tag{2}$$

where  $\mathbf{P}'_0 = \frac{d}{d\beta} \mathbf{P}_\beta |_{\beta=0}$  and  $\mathbf{c} = \sum_{k=0}^{\infty} \mathbf{P}_0^k (2\mathbf{x} - \mathbf{1})$  is the accumulated average abundance difference during the whole evolution at the neutral drift (see detailed calculations in Methods). In fact, condition (2) holds for a large class of evolutionary dynamics which admit a unique limiting stationary distribution and an equal abundance of strategy  $A$  and  $B$  at the neutral drift. For example, the death-birth, birth-death, and pairwise comparison process with symmetric mutations all belong to this class<sup>11,12,24,28,30,45</sup>.

For aspiration-based update rules, individuals update their strategies independently when  $\beta = 0$ . This makes it possible for us to calculate the exact formula of  $\mathbf{c}$  even if each individual uses distinct update functions. In a nutshell, we transform the calculation under the original  $N$ -dimensional Markov chain with  $2^N$  states to that under  $N$  one-dimensional Markov chains with 2 states by virtue of the independence of strategy updating (see detailed calculations in Supplementary Note 3.2). In addition, we adapt our method for imitation-based update rules with a shared update function, which is shown to be equivalent to that in ref. <sup>46</sup>. In particular, we give the condition for strategy success on any weighted graphs under pairwise comparison rules (see Supplementary Note 3.3 for detailed calculations).

In the following, we mainly focus on aspiration-based update rules and explore how symmetric or asymmetric aspirations affect the evolutionary outcomes.

**Personalized and symmetric aspirations.** Let us first consider personalized and symmetric aspirations. Under weak selection,

we find that strategy  $A$  prevails over  $B$  if

$$a + b > c + d, \quad (3)$$

and strategy  $B$  prevails over  $A$  if  $a + b < c + d$ . This result holds for all weighted graphs without self-loops, for all distributions of aspirations, and for arbitrary number of update functions. Furthermore, if strategy  $A$  and  $B$  are both best replies to themselves (i.e.,  $a > c$  and  $b < d$ ), our result reduces to the classical concept of risk-dominance. It indicates that under the limit of weak selection, aspiration dynamics always select the risk-dominant strategy, which has a larger basin of attraction. In Fig. 2, we set payoff values of the game with  $b = 0$ ,  $c = 5$ , and  $d = 1$ . Under this game, condition (3) predicts that for any weighted graphs without self-loops,  $a > 6$  indicates strategy  $A$  prevails over  $B$  (equivalently,  $\langle x_A \rangle > 1/2$ );  $a < 6$  leads to  $\langle x_A \rangle < 1/2$ . Our simulation results in Fig. 2 match the theoretical predictions perfectly.

The above condition significantly generalizes previous results: the selection of risk-dominant strategy on unweighted regular graphs (where all the individuals have the same number of neighbors and each individual interacts with their neighbors under identical rates)<sup>31–33</sup> is generalized to non-regular and weighted graphs with individualized update functions and personalized aspirations.

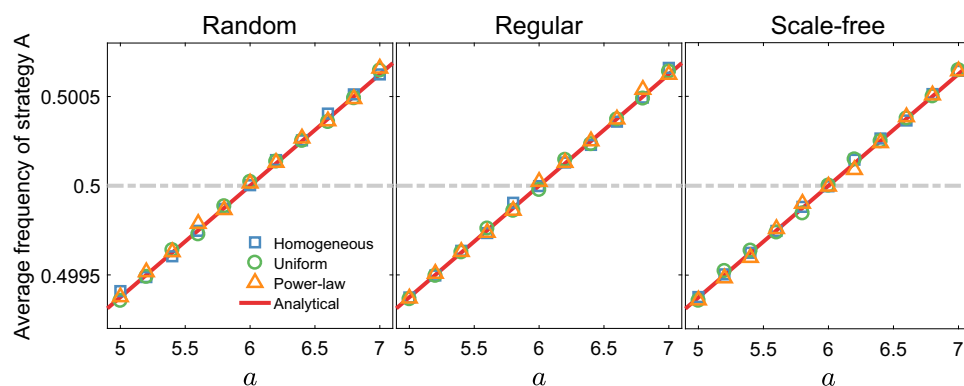
For an intuitive understanding of our result, we offer the following explanations. Under weak selection, the expected payoffs of playing strategy  $A$  and  $B$  are evaluated at the neutral ( $\beta = 0$ ) stationary distribution of the aspiration dynamics (see Methods for details). In this distribution, individuals update strategies independently, which makes their strategies uncorrelated. Individual  $l$  thus on average interacts with neighbors using strategy  $A$  as many times as those using  $B$ . This means that the expected payoffs of  $l$  are  $\pi_{l,A} = (1/2)(a + b)$  and  $\pi_{l,B} = (1/2)(c + d)$  when it plays strategy  $A$  and  $B$ , respectively. If  $\pi_{l,A} > \pi_{l,B}$ , individual  $l$  is more satisfied when it uses strategy  $A$  and the switching rate from  $A$  to  $B$  is less than that from  $B$  to  $A$ . Note that  $\pi_{l,A} > \pi_{l,B}$  is equivalent to  $a + b > c + d$ . Therefore, individual  $l$  is more likely to be an  $A$ -player if  $a + b > c + d$ . Since the above logic applies to any individual, the condition  $a + b > c + d$  actually makes all the individuals feel more satisfied when they play strategy  $A$ . As a consequence, the average frequency of  $A$ -players in the population is greater than that of  $B$ -players. Similarly,  $a + b < c + d$  results in more satisfaction when

individuals play strategy  $B$ , which makes the average frequency of  $A$ -players less than that of  $B$ -players.

Comparing our result with that in well-mixed populations (equivalent to a complete graph in our model)<sup>31</sup>, we show that population structure does not alter the condition for strategy success. In other words, the condition for strategy success under aspiration dynamics is robust to the underlying population structure. The robustness property has practical advantages on strategy selection<sup>11</sup>: (i) for a fixed game, the predictions are the same for a large class of population structures; (ii) to tell which strategy succeeds, the population can be assumed to be well-mixed.

Besides, by generalizing the main Theorem in ref. <sup>32</sup> to non-regular graphs, our result is related to the structure coefficient  $\sigma$ <sup>8,11</sup> derived for weak selection. It is shown that  $\sigma$  depends on the update rule and the population structure (including the population size  $N$ ). But it does not depend on the payoff entries. It summarizes the effect of population structure on the condition for strategy success. Intuitively, for strategy  $A$  to be favored over strategy  $B$ ,  $\sigma$  quantifies the required degree of assortment among individuals who use the same strategy<sup>11</sup>, and  $\sigma > 1$  implies that individuals with the same strategy are more likely to interact with each other than those with different strategies. For aspiration dynamics, we prove that  $\sigma = 1$  for a large class of population structures and there is no dependence on the population size. This contrasts with the result obtained under imitation-based rules, which are shown to sensitively depend on the population structure and the population size<sup>11,24</sup>. To better illustrate the difference between aspiration-based and imitation-based rules, we plot Fig. 3 to compare the structure coefficients on three common graphs. Here,  $\sigma = 1$  indicates that self-evaluation-based rules do not lead to assortment of strategies for the purpose of strategy selection.

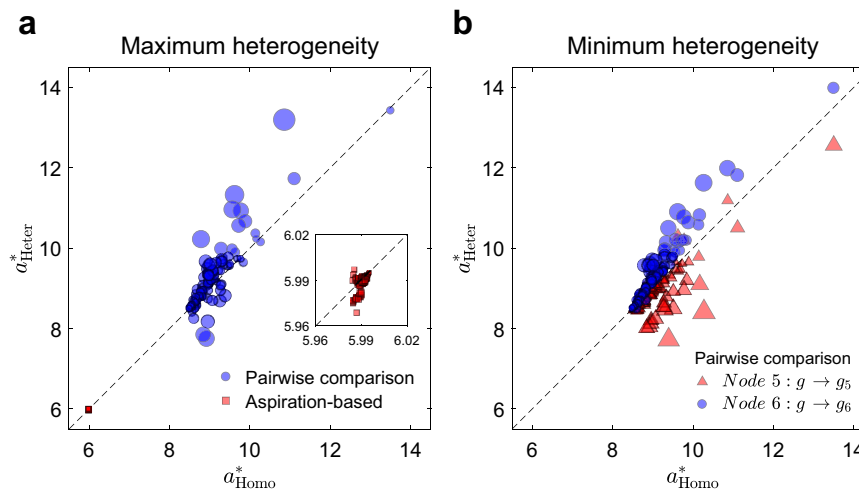
In the meanwhile, we show that the results under imitation-based rules are also sensitive to the heterogeneity of update functions, even if we consider minimum heterogeneity (see Fig. 4b and Supplementary Fig. 1). In contrast, maximum heterogeneity of update functions does not alter the evolutionary outcomes induced by aspiration-based rules (see Fig. 4a). Such results demonstrate that aspiration-based rules lead to evolutionary outcomes robust to individual heterogeneities while imitation-based rules show great sensitivity.



**Fig. 2 Robust predictions generated by aspiration dynamics on weighted networks.** For the game, we set payoff value  $b = 0$ ,  $c = 5$ ,  $d = 1$  and leave  $a$  as a tunable parameter. We plot the average frequency of strategy  $A$ ,  $\langle x_A \rangle$ , as a function of  $a$ . Symbols represent simulation results while solid lines are analytical ones. We construct weighted graphs by first generating an undirected graph with average degree  $\bar{k}$  and then assigning weights to edges. The undirected graphs considered are random graph (left), regular graph (middle), and scale-free network (right). For each type of network, we test three edge weight distributions: homogeneous—every edge has weight one (i.e., unweighted network); uniform—edge weights are uniformly selected from the integer set  $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ ; power-law—edge weights are randomly selected from the discrete power-law distribution (Zipf distribution with the value of the exponent equal to 3). Each data point is obtained by averaging  $\langle x_A \rangle$  in 200 independent runs. For each run, we calculate  $\langle x_A \rangle$  by averaging the frequency of strategy  $A$  in the last  $1 \times 10^7$  time steps after a transient time of  $1 \times 10^7$  time steps. Other parameters:  $N = 1000$ ,  $\bar{k} = 6$ ,  $\alpha_l = 2.0$  ( $l = 1, 2, \dots, N$ ), and  $\beta = 0.01$ .

Update rule	Well-mixed	Regular graph	Star
Aspiration-based	1	1	1
Pairwise comparison	$\frac{N - 2}{N}$	$\frac{N - 2}{N}$	$\frac{N - 2}{2(N - 1)}$
Death-birth	$\frac{N - 2}{N}$	$\frac{(k + 1)N - 4k}{(k - 1)N}$	1

**Fig. 3 The structure coefficients  $\sigma$  for aspiration-based and imitation-based update rules.** The two imitation-based rules shown here are pairwise comparison<sup>29,30</sup> and death-birth update rule<sup>28</sup>. Here, the parameter  $N$  represents the population size and  $k$  the degree of the regular graph (i.e., the number of neighbors each individual has). For strategy  $A$  to be favored over strategy  $B$ , the structure coefficient  $\sigma$  can be interpreted as the required degree of assortment among individuals who use the same strategy<sup>11</sup>. All the  $\sigma$ s in the table are obtained under the limit of weak selection. In addition, the limit of rare mutation is assumed under imitation-based rules. We also derive a general formula under pairwise comparison update rules for any weighted graphs in Supplementary Note 3.3.



**Fig. 4 Robustness of aspiration dynamics vs. sensitivity of imitative dynamics to heterogeneity of decision-making functions for all networks of size six ( $N = 6$ ).** There are 112 connected and unweighted networks of size six. For each network, we calculate the critical value of  $a$ ,  $a^*$ , above which  $\langle x_A \rangle > \langle x_B \rangle$  and below which  $\langle x_A \rangle < \langle x_B \rangle$ . The critical value  $a^*$  under two populations are calculated: (i)  $a_{\text{Homo}}^*$ , a homogeneous population where everyone shares the same decision-making function  $g(u)$ ; (ii)  $a_{\text{Heter}}^*$ , a heterogeneous population where each individual has its own decision-making functions, i.e., individual  $i$  uses  $g_i(u)$ . In panels (a) and (b), each symbol represents a pair of critical values ( $a_{\text{Homo}}^*$ ,  $a_{\text{Heter}}^*$ ) calculated under one of the networks. The size of the symbols indicates how much the heterogeneity of decision-making functions affects the evolutionary outcomes, which is proportional to the absolute value of  $a_{\text{Heter}}^* - a_{\text{Homo}}^*$ . The larger the size of the symbol, the more sensitive the evolutionary outcome to individual heterogeneity. In panel a, we consider maximum heterogeneity of decision-making functions (i.e., for any  $i \neq j$ ,  $g_i \neq g_j$ ) under both aspiration-based (red squares) and imitation-based (blue circles) rules. In panel b, we consider the minimum heterogeneity under imitation-based rules, where all the individuals share the same decision-making function  $g(u)$  except one. Here, the blue circles are the results obtained when  $g_6 \neq g$ , and the red triangles are the results when  $g_5 \neq g$ . In all the calculations, we set  $g(u) = g_1(u) = 1/(1 + \exp(-u))$ ,  $g_2(u) = (1 + \text{erf}(u))/2$ ,  $g_3(u) = (1 + \tanh(u))/2$ ,  $g_4(u) = 1/(1 + \exp(-u/2))$ ,  $g_5(u) = 1/(1 + 10 \exp(-u))$ , and  $g_6(u) = 10/(10 + \exp(-u))$ . Other parameters:  $b = 0$ ,  $c = 5$ ,  $d = 1$ ,  $\alpha_l = 1.0$  ( $l = 1, 2, \dots, N$ ),  $\beta = 0.01$ , and  $\mu \rightarrow 0$  (see Supplementary Note 3.3 for definitions).

**Asymmetric aspirations contingent on strategies.** When individuals' aspirations are contingent on their strategy in use (and thereby asymmetric), we find that the condition for strategy  $A$  to be favored over  $B$  under weak selection is  $a + b > c + d - 2(\alpha_B - \alpha_A)$ , where  $\alpha_X$  is the aspiration of  $X$ -

players ( $X = A, B$ ). Note that the condition now depends on aspirations. Nonetheless, it is still robust to population structures, which generates invariant predictions for a large class of population structures. Intuitively, the symmetry breaking of aspiration levels leads to additional asymmetry between strategy  $A$  and  $B$ :

A-players not only gain a different payoff but also have a different benchmark for satisfaction from B-players. The expected level of satisfaction is now modified as  $(1/2)(a + b) - \alpha_A$  and  $(1/2)(c + d) - \alpha_B$  for A-players and B-players, respectively. This modification alters the condition of risk-dominance derived under personalized aspirations and results in the dependence on aspiration levels.

So far, we only consider average payoffs. Our framework also applies to accumulated payoffs (see Supplementary Note 3.2.4 for the condition for strategy success). We show that the condition of risk-dominance is invariant under accumulated payoffs, provided that aspirations are not contingent on strategies and the selection intensity is weak. In addition, we also verify the condition for strategy success (i.e., inequality (3)) under the classical payoff scheme of  $R = 1, S, T, P = 0$  (see Supplementary Fig. 2), and confirm the robustness of aspiration dynamics as well as the sensitivity of imitative dynamics to the heterogeneity of decision-making functions on large networks (see Supplementary Fig. 3).

## Discussion

In this work, we present a general framework to study aspiration dynamics in heterogeneous populations, which makes it possible to study the joint effect of heterogeneous population structure, personalized aspiration values, and individualized update rules on the evolutionary outcomes. Previous studies<sup>23,31–33</sup>, due to the limitation of their methodology, can only handle regular graphs. Under our framework, we show that under weak selection, the condition for one strategy to be selected over the other is invariant on different population structures and under various kinds of individual heterogeneities. Moreover, this condition coincides with the condition of risk-dominance. It indicates that aspiration dynamics always select the risk-dominant strategy. When individuals' aspirations are contingent on strategies and thus asymmetric, the condition for strategy success is altered and determined by the difference between the aspirations of distinct strategies. In this case, cooperation can evolve in the Prisoner's Dilemma if individuals aspire more when they defect.

Our framework can also be used to study imitation-based update rules. We demonstrate that our approach is equivalent to that in refs.<sup>24,46</sup> when all the individuals share the same imitation-based update function (see Supplementary Note 3.3 for details). Moreover, our results for the pairwise comparison rule show that cooperation can never evolve on any weighted networks since the critical benefit-to-cost ratio is negative and the evolutionary outcomes are greatly affected by the heterogeneity of update functions. This confirms that evolutionary outcomes induced by imitation-based rules are very sensitive to model details. Compared with the results under aspiration dynamics, it highlights the advantage of aspiration-based update rules, which generates robust evolutionary outcomes.

In a nutshell, the primary contribution of our work is three-fold: (i) we derive a general condition for strategy success that applies to a large class of evolutionary dynamics (see Supplementary Note 3 for details); (ii) we prove that for aspiration dynamics, the condition for strategy success is invariant on any weighted networks, which reveals the remarkable robustness of aspiration dynamics to the underlying population structure; (iii) the robustness property of aspiration dynamics is shown to be universal with respect to various kinds of heterogeneities and their aggregations, including heterogeneity of aspiration values, heterogeneity of update functions, and heterogeneity of social ties.

Although our work provides a general framework to study evolutionary dynamics in heterogeneous populations, theoretical results are obtained under the limit of weak selection. In this selection regime, the payoffs garnered by individuals affect

minimally their probability of changing strategies. If the selection intensity becomes strong, individuals' strategy updating will be strongly affected by payoffs and aspirations. In this case, our theoretical results may no longer apply. Despite this limitation, it is still necessary and useful to conduct theoretical analysis under weak selection since (i) it may be by far the only way to obtain analytical results for evolutionary dynamics on heterogeneous networks<sup>24,47</sup>, and (ii) the theoretical predictions obtained can provide guidance for future studies (e.g., testing such predictions experimentally<sup>15–18</sup>).

For self-evaluation-based update rules we focus on in this paper, in addition to the irrelevance of social information, they also have other features: self-evaluation-based rules are innovative<sup>7</sup>, which means they can revive strategies absent in the neighborhood without additional mechanisms such as random exploration or mutation; they prescribe increasing tendency to cooperate when more cooperators are present in the neighborhood for the Prisoner's Dilemma (similar to conditional cooperators<sup>48</sup>). These features seem to be consistent with the recent findings on the possible features of human strategy updating<sup>15–18</sup>. This suggests that self-evaluation-based rules may be a good candidate for human strategy updating, which needs further empirical test.

In addition, aspiration-based self-evaluation rules are related to reinforcement learning. The rationale behind reinforcement learning is the law of effect stated by Thorndike in 1898: actions bringing satisfactory effect will be more likely to be repeated and those leading to discomfort will be less likely to occur. This is similar to our stochastic update rules, except the reinforcement of actions<sup>49,50</sup>. In practice, aspiration can also evolve based on past experience<sup>22</sup>. Although these features are not considered in our models, our work provides an important step towards multi-agent learning in heterogeneous populations, whereas literature on reinforcement learning usually focuses on the simplest two-person repeated games (see a few recent exceptions on regular graphs<sup>51,52</sup>). Extending our model to incorporate aspiration adaptation and reinforcement of actions is a future direction.

For the evolution of human cooperation, our work suggests the investigations of which update rules human actually uses for strategy updating. A promising direction is to conduct experiments explicitly manipulating the information availability or monitoring the information request during the game<sup>14,53</sup>. Then, based on the distinct informational requirements of self-evaluation-based and imitation-based rules, we may infer under what conditions human subjects tend to use these two classes of update rules and how they implement them. For theoretical studies, our work reveals a class of update rules which generate robust predictions for strategy success on a large class of population structures. The reason may lie in the irrelevance of social peers' payoffs for strategy updating. It remains unclear what other assumptions in update rules crucially affect evolutionary outcomes. Future work along this line may lead to a deeper understanding of how update rules alter the evolutionary outcomes, which may help design the optimal decision-making rules for cooperation.

## Methods

**Notation.** The population consists of  $N$  individuals. Each individual either uses strategy A or strategy B. We use  $s_i$  to denote the strategy of individual  $i$ :  $s_i = 1$  if individual  $i$  uses strategy A, otherwise  $s_i = 0$ . The state of the population is represented by a column vector  $\mathbf{s} = (s_1, s_2, s_3, \dots, s_N)^T$  and  $\mathbf{s} = \mathbf{1}$  means all the individuals in the population use strategy A. At state  $\mathbf{s}$ , we denote the frequency (i.e., abundance) of strategy A as  $x_A(\mathbf{s}) = \sum_{i=1}^N s_i / N$  and that of strategy B as  $x_B(\mathbf{s}) = 1 - x_A(\mathbf{s})$ . Since each individual can use strategy A or strategy B, the number of all possible states of the population is  $M = 2^N$ . A convenient way of indexing the state is to convert the binary vector  $\mathbf{s}$  to a decimal number and plus one, which makes the index of states range from 1 to  $M$ . Here, we use the column vector

$\mathbf{x} = (x_A(s_1), x_A(s_2), \dots, x_A(s_M))$  to represent the frequency of strategy  $A$  at each state.

Meanwhile, individuals play game (1) with their neighbors and obtain edge-weighted average payoffs. For instance, at state  $\mathbf{s}$ , individual  $l$  gets a payoff  $\pi_l(\mathbf{s}) = \sum_{k=1}^N w_{lk}[as_l s_k + bs_l(1 - s_k) + c(1 - s_l)s_k + d(1 - s_l)(1 - s_k)]/d_l$ .

**General condition for strategy success.** For evolutionary dynamics which can be modeled by an irreducible and aperiodic Markov chain, a unique limiting stationary distribution  $\mathbf{u} = (u_1, u_2, \dots, u_j, \dots, u_M)^T$  is guaranteed, where  $u_j$  means the probability to occur at state  $j$  in the stationary distribution. We denote the average frequency of strategy  $A$  in the stationary distribution as  $\langle x_A \rangle = \mathbf{u}^T \mathbf{x}$  and that of  $B$  as  $\langle x_B \rangle = \mathbf{u}^T (\mathbf{I} - \mathbf{x})$ . In addition, a transition matrix  $\mathbf{P}$  is constructed where its  $(i, j)$ -th element  $p_{ij}$  represents the probability to transit from state  $i$  to state  $j$ . Since the stationary distribution  $\mathbf{u}$ , transition matrix  $\mathbf{P}$ , and average frequency of strategies  $\langle x_A \rangle$ ,  $\langle x_B \rangle$  are determined by the selection intensity  $\beta$ , we rewrite them explicitly as  $\mathbf{u}_\beta$ ,  $\mathbf{P}_\beta$ ,  $\langle x_A \rangle_\beta$ , and  $\langle x_B \rangle_\beta$ . By the definition of stationary distribution,  $\mathbf{u}_\beta^T \mathbf{P}_\beta = \mathbf{u}_\beta^T$ . Differentiating both sides with respect to  $\beta$  at  $\beta = 0$  and rearranging the items, we have

$$(\mathbf{u}_0^T)^T = \mathbf{u}_0^T \mathbf{P}'_0 (\mathbf{I} - \mathbf{P}_0 + \mathbf{1u}_0^T)^{-1} \tag{4}$$

where  $\mathbf{I}$  is the identity matrix of dimension  $M$ ,  $\mathbf{u}'_0 = \frac{d}{d\beta} \mathbf{u}_\beta|_{\beta=0}$ , and  $\mathbf{P}'_0 = \frac{d}{d\beta} \mathbf{P}_\beta|_{\beta=0}$ . Under weak selection  $\beta \rightarrow 0$ ,

$$\langle x_A \rangle_\beta - \langle x_B \rangle_\beta = \mathbf{u}_0^T (2\mathbf{x} - \mathbf{1}) + (\mathbf{u}'_0)^T (2\mathbf{x} - \mathbf{1})\beta + O(\beta^2). \tag{5}$$

Here, we assume that for the evolutionary dynamics we are focusing on, the average frequency of strategy  $A$  equals to that of strategy  $B$  at the neutral drift  $\beta = 0$ , which means that  $\mathbf{u}_0^T (2\mathbf{x} - \mathbf{1}) = 0$ . Therefore, under weak selection, strategy  $A$  prevails over  $B$  (i.e.,  $\langle x_A \rangle_\beta > \langle x_B \rangle_\beta$ ) if

$$\mathbf{u}_0^T \mathbf{P}'_0 \sum_{k=0}^{\infty} \mathbf{P}_0^k (2\mathbf{x} - \mathbf{1}) > 0, \tag{6}$$

and strategy  $B$  prevails over  $A$  if  $\mathbf{u}_0^T \mathbf{P}'_0 \sum_{k=0}^{\infty} \mathbf{P}_0^k (2\mathbf{x} - \mathbf{1}) < 0$ . Let  $\mathbf{c} = \sum_{k=0}^{\infty} \mathbf{P}_0^k (2\mathbf{x} - \mathbf{1}) = (c_1, c_2, \dots, c_M)^T$ , which represents the accumulated average abundance difference between strategy  $A$  and  $B$  during the whole evolution. Then, we get condition (2) in the main text.

**Condition for strategy success under aspiration dynamics.** For aspiration-based update rules, we first define individual  $l$ 's update function as  $g_l(u)$  ( $l = 1, 2, \dots, N$ ) and these functions represent the tendency to switch strategies. Here,  $u = \beta(\alpha_l - \pi_l)$ , where  $\alpha_l$  is individual  $l$ 's aspiration level,  $\pi_l$  is its payoff, and  $\beta > 0$  is the intensity of selection<sup>24,36</sup>. Weak selection means  $\beta \ll 1$  and  $\beta = 0$  is the neutral drift<sup>11</sup>. In addition, each function  $g_l(u)$  should satisfy (i) it is a probability, i.e.,  $g_l(u) \in [0, 1]$  for  $u \in \mathbb{R}$ ; (ii) it is a strictly increasing function of  $u$ , i.e.,  $g'_l(u) = dg_l(u)/du > 0$  for all  $u$ , which indicates that individuals with higher payoffs should have a lower tendency to switch; (iii)  $g_l(0) > 0$ , which avoids frozen dynamics at the neutral drift.

The population dynamics governed by aspiration-based update rules can be described by an irreducible and aperiodic Markov chain<sup>32</sup>. Meanwhile, such dynamics also lead to an equal average abundance of strategy  $A$  and  $B$  at the neutral drift. Therefore, the general condition (2) applies to aspiration-based update rules.

To calculate the exact formula of  $\mathbf{c}$  in condition (2), we utilize the fact that individuals update their strategies independently when  $\beta = 0$  for aspiration-based update rules. This makes it possible to calculate  $\mathbf{c}$  by summing up all the individual contributions. For each individual, its dynamics can be modeled by a one-dimensional Markov chain with 2 states. For instance, the transition matrix for individual  $l$  is

$$\begin{pmatrix} 1 & 0 \\ 1 - \frac{1}{N}g_l(0) & \frac{1}{N}g_l(0) \\ 0 & \frac{1}{N}g_l(0) & 1 - \frac{1}{N}g_l(0) \end{pmatrix} \tag{7}$$

and its individual contribution to the accumulated average abundance difference during the whole evolution is  $(2s_l(l) - 1)/(2g_l(0))$  where  $s_l(l)$  is the strategy of individual  $l$  at state  $i$ . Based on these, we have

$$c_i = \sum_{l=1}^N \frac{2s_l(l) - 1}{2g_l(0)}. \tag{8}$$

Denote  $h(\mathbf{s}_i) = \sum_{j=1}^M p'_{ij} c_j = \sum_{j=1}^M p'_{ij} \sum_{l=1}^N \frac{2s_l(l) - 1}{2g_l(0)}$ . Condition (2) in the main text implies that strategy  $A$  prevails over  $B$  if

$$\langle h(\mathbf{s}) \rangle_0 > 0, \tag{9}$$

where the bracket  $\langle \cdot \rangle_0$  means to take the average over the neutral stationary distribution (i.e., when  $\beta = 0$ ). To evaluate  $\langle h(\mathbf{s}) \rangle_0$ , we need to know the correlation of strategies in the neutral stationary distribution, i.e.,  $\langle s_l s_k \rangle_0$  for any  $l, k$ . Under aspiration dynamics, at the neutral drift  $\beta = 0$ , individuals' strategy updating does

not depend on the aspiration level, the payoff, and the population structure. This implies that the transition probabilities between any two states are thus the same in both directions, which indicates  $\mathbf{P}_0 = \mathbf{P}_0^T$ . By the uniqueness of the stationary distribution  $\mathbf{u}_0$  and the property of the transition (stochastic) matrix  $\mathbf{P}_0 \mathbf{1} = \mathbf{1}$ , we have  $\mathbf{u}_0 = 2^{-N} \mathbf{1}$ . This leads to that at the neutral stationary distribution, each individual plays strategy  $A$  with probability one-half, i.e.,  $\langle s_l \rangle_0 = 1/2$  for all  $l$ . Furthermore, due to the independence of individuals' strategy updating,  $\langle s_l s_k \rangle_0 = \langle s_l \rangle_0 \langle s_k \rangle_0 = 1/4$  when  $l \neq k$ . Based on these, we have that the correlations of strategies at the neutral stationary distribution are

$$\langle s_l s_k \rangle_0 = \frac{1 + \delta_{lk}}{4}, \tag{10}$$

where  $\delta_{lk} = 0$  if  $l \neq k$  and  $\delta_{lk} = 1$  if  $l = k$ . Equation (10) indicates that aspiration dynamics do not lead to assortment of strategies in the neutral stationary distribution.

**Personalized and symmetric aspirations.** For personalized and symmetric aspirations, each individual  $l$  has its own aspiration  $\alpha_l$  ( $l = 1, 2, \dots, N$ ) and this aspiration does not depend on the strategy of individual  $l$ . In this case, we have

$$h(\mathbf{s}) = \frac{1}{N} \sum_{l=1}^N \frac{g'_l(0)}{g_l(0)} [(1 - 2s_l)\alpha_l + (1 - 2s_l)\pi_l(\mathbf{s})], \tag{11}$$

and

$$\langle h(\mathbf{s}) \rangle_0 = \frac{1}{4N} \left( \sum_{l=1}^N \frac{g'_l(0)}{g_l(0)} \right) (a + b - c - d). \tag{12}$$

Since  $g'_l(0) > 0$  and  $g_l(0) > 0$  for any  $l$ , equation (12) implies that for symmetric aspirations, strategy  $A$  prevails over  $B$  if  $a + b > c + d$  and strategy  $B$  prevails over  $A$  if  $a + b < c + d$ . Note that the condition for strategy success does not depend on the population structure at all, which highlights the intrinsic difference between the evolutionary dynamics induced by aspiration-based and imitation-based update rules.

**Asymmetric aspirations contingent on strategies.** For aspirations contingent on the strategy in use, individuals playing strategy  $A$  have aspiration level  $\alpha_A$  while those using  $B$  have  $\alpha_B$ . Since coefficients  $\mathbf{c}$  is evaluated at  $\beta = 0$ , the asymmetry of aspirations thus does not affect  $\mathbf{c}$ . However,  $h(\mathbf{s})$  now depends on both  $\alpha_A$  and  $\alpha_B$ , and it is modified as

$$h(\mathbf{s}) = \frac{1}{N} \sum_{l=1}^N \frac{g'_l(0)}{g_l(0)} [(1 - s_l)\alpha_B - s_l\alpha_A + (1 - 2s_l)\pi_l(\mathbf{s})]. \tag{13}$$

Meanwhile, at the neutral stationary distribution, strategy correlations  $\langle s_l s_k \rangle_0$  are also independent of aspirations. Based on these, we have

$$\langle h(\mathbf{s}) \rangle_0 = \frac{1}{4N} \left( \sum_{l=1}^N \frac{g'_l(0)}{g_l(0)} \right) (a + b - c - d + 2\alpha_B - 2\alpha_A), \tag{14}$$

which implies under asymmetric aspirations, strategy  $A$  prevails over  $B$  if

$$a + b > c + d - 2(\alpha_B - \alpha_A),$$

and strategy  $B$  prevails over  $A$  if  $a + b < c + d - 2(\alpha_B - \alpha_A)$ .

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

All data generated or analyzed during this study are included within the paper and its supplementary information files.

**Code availability**

We use Microsoft Visual Studio Community 2019 and Matlab 2020a to write the custom Fortran code and analyze the data, respectively. The computer codes for Monte Carlo simulations used in this work are available in Zenodo with the identifier <https://doi.org/10.5281/zenodo.4723407>.

Received: 7 April 2020; Accepted: 5 May 2021; Published online: 31 May 2021

**References**

- Hamilton, W. D. The genetical evolution of social behaviour. i. *J. Theor. Biol.* 7, 1–16 (1964).
- Hamilton, W. D. The genetical evolution of social behaviour. ii. *J. Theor. Biol.* 7, 17–52 (1964).

3. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
4. Nowak, M. A. *Evolutionary Dynamics* (Harvard Univ. Press, 2006).
5. Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359**, 826–829 (1992).
6. Hauert, C. & Doebeli, M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **428**, 643 (2004).
7. Szabó, G. & Fáth, G. Evolutionary games on graphs. *Phys. Rep.* **446**, 97–216 (2007).
8. Nowak, M. A., Tarnita, C. E. & Antal, T. Evolutionary dynamics in structured populations. *Philos. Trans. R. Soc. B* **365**, 19–30 (2010).
9. Débarre, F., Hauert, C. & Doebeli, M. Social evolution in structured populations. *Nat. Commun.* **5**, 3409 (2014).
10. Roca, C. P., Cuesta, J. A. & Sánchez, A. Evolutionary game theory: temporal and spatial effects beyond replicator dynamics. *Phys. Life Rev.* **6**, 208–249 (2009).
11. Tarnita, C. E., Ohtsuki, H., Antal, T., Fu, F. & Nowak, M. A. Strategy selection in structured populations. *J. Theor. Biol.* **259**, 570–581 (2009).
12. Allen, B. & Nowak, M. A. Games on graphs. *EMS Surv. Math. Sci.* **1**, 113–151 (2014).
13. Molleman, L., van den Berg, P. & Weissing, F. J. Consistent individual differences in human social learning strategies. *Nat. Commun.* **5**, 3570 (2014).
14. van den Berg, P., Molleman, L. & Weissing, F. J. Focus on the success of others leads to selfish behavior. *Proc. Natl Acad. Sci. USA* **112**, 2912–2917 (2015).
15. Grujić, J., Fosco, C., Araujo, L., Cuesta, J. A. & Sánchez, A. Social experiments in the mesoscale: humans playing a spatial prisoner's dilemma. *PLoS ONE* **5**, e13749 (2010).
16. Gracia-Lázaro, C. et al. Heterogeneous networks do not promote cooperation when humans play a prisoner's dilemma. *Proc. Natl Acad. Sci. USA* **109**, 12922–12926 (2012).
17. Grujić, J. et al. A comparative analysis of spatial prisoner's dilemma experiments: conditional cooperation and payoff irrelevance. *Sci. Rep.* **4**, 4615 (2014).
18. Rand, D. G., Nowak, M. A., Fowler, J. H. & Christakis, N. A. Static network structure can stabilize human cooperation. *Proc. Natl Acad. Sci. USA* **111**, 17093–17098 (2014).
19. Karandikar, R., Mookherjee, D., Ray, D. & Vega-Redondo, F. Evolving aspirations and cooperation. *J. Econ. Theory* **80**, 292–331 (1998).
20. Posch, M., Pichler, A. & Sigmund, K. The efficiency of adapting aspiration levels. *Proc. R. Soc. Lond. B* **266**, 1427–1435 (1999).
21. Chen, X. & Wang, L. Promotion of cooperation induced by appropriate payoff aspirations in a small-world networked game. *Phys. Rev. E* **77**, 017103 (2008).
22. Bendor, J., Diermeier, D., Siegel, D. A. & Ting, M. M. *A Behavioral Theory of Elections* (Princeton Univ. Press, 2011).
23. Du, J., Wu, B., Altrock, P. M. & Wang, L. Aspiration dynamics of multi-player games in finite populations. *J. R. Soc. Interface* **11**, 20140077 (2014).
24. Allen, B. et al. Evolutionary dynamics on any population structure. *Nature* **544**, 227–230 (2017).
25. Ohtsuki, H. & Nowak, M. A. The replicator equation on graphs. *J. Theor. Biol.* **243**, 86–97 (2006).
26. Cardillo, A., Gómez-Gardeñes, J., Vilone, D. & Sánchez, A. Co-evolution of strategies and update rules in the prisoner's dilemma game on complex networks. *New J. Phys.* **12**, 103034 (2010).
27. Zukewich, J., Kurella, V., Doebeli, M. & Hauert, C. Consolidating birth-death and death-birth processes in structured populations. *PLoS ONE* **8**, e54639 (2013).
28. Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M. A. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502–505 (2006).
29. Szabó, G. & Tóke, C. Evolutionary prisoner's dilemma game on a square lattice. *Phys. Rev. E* **58**, 69–73 (1998).
30. Traulsen, A., Nowak, M. A. & Pacheco, J. M. Stochastic dynamics of invasion and fixation. *Phys. Rev. E* **74**, 011909 (2006).
31. Du, J., Wu, B. & Wang, L. Aspiration dynamics in structured population acts as if in a well-mixed one. *Sci. Rep.* **5**, 8014 (2015).
32. Wu, B. & Zhou, L. Individualised aspiration dynamics: calculation by proofs. *PLoS Comput. Biol.* **14**, e1006035 (2018).
33. Zhou, L., Wu, B., Vasconcelos, V. V. & Wang, L. Simple property of heterogeneous aspiration dynamics: beyond weak selection. *Phys. Rev. E* **98**, 062124 (2018).
34. Simon, H. A. Theories of decision-making in economics and behavioral science. *Am. Econ. Rev.* **49**, 253–283 (1959).
35. Wu, B., García, J., Hauert, C. & Traulsen, A. Extrapolating weak selection in evolutionary games. *PLoS Comput. Biol.* **9**, e1003381 (2013).
36. Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).
37. Traulsen, A., Pacheco, J. M. & Nowak, M. A. Pairwise comparison and selection temperature in evolutionary game dynamics. *J. Theor. Biol.* **246**, 522–529 (2007).
38. Antal, T., Traulsen, A., Ohtsuki, H., Tarnita, C. E. & Nowak, M. A. Mutation-selection equilibrium in games with multiple strategies. *J. Theor. Biol.* **258**, 614–622 (2009).
39. Fu, F., Wang, L., Nowak, M. A. & Hauert, C. Evolutionary dynamics on graphs: efficient method for weak selection. *Phys. Rev. E* **79**, 046707 (2009).
40. Wu, B., Altrock, P. M., Wang, L. & Traulsen, A. Universality of weak selection. *Phys. Rev. E* **82**, 046106 (2010).
41. Tarnita, C. E., Wage, N. & Nowak, M. A. Multiple strategies in structured populations. *Proc. Natl Acad. Sci. USA* **108**, 2334–2337 (2011).
42. Bentley, R. A., Hahn, M. W. & Shennan, S. J. Random drift and culture change. *Proc. R. Soc. Lond. B* **271**, 1443–1450 (2004).
43. Bentley, R. A., Lipo, C. P., Herzog, H. A. & Hahn, M. W. Regular rates of popular culture change reflect random copying. *Evol. Hum. Behav.* **28**, 151–158 (2007).
44. Traulsen, A., Nowak, M. A. & Pacheco, J. M. Stochastic payoff evaluation increases the temperature of selection. *J. Theor. Biol.* **244**, 349–356 (2007).
45. Vasconcelos, V. V., Santos, F. P., Santos, F. C. & Pacheco, J. M. Stochastic dynamics through hierarchically embedded Markov chains. *Phys. Rev. Lett.* **118**, 058301 (2017).
46. Allen, B. & McAvoy, A. A mathematical formalism for natural selection with arbitrary spatial and genetic structure. *J. Math. Biol.* **78**, 1147–1210 (2019).
47. Ibsen-Jensen, R., Chatterjee, K. & Nowak, M. A. Computational complexity of ecological and evolutionary spatial dynamics. *Proc. Natl Acad. Sci. USA* **112**, 15636–15641 (2015).
48. Fischbacher, U., Gächter, S. & Fehr, E. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404 (2001).
49. Börgers, T. & Sarin, R. Learning through reinforcement and replicator dynamics. *J. Econ. Theory* **77**, 1–14 (1997).
50. Erev, I. & Roth, A. E. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* **88**, 848–881 (1998).
51. Ezaki, T., Horita, Y., Takezawa, M. & Masuda, N. Reinforcement learning explains conditional cooperation and its moody cousin. *PLoS Comput. Biol.* **12**, e1005034 (2016).
52. Ezaki, T. & Masuda, N. Reinforcement learning account of network reciprocity. *PLoS ONE* **12**, e0189220 (2017).
53. Kovářík, J., Mengel, F. & Romero, J. G. Learning in network games. *Quant. Econ.* **9**, 85–139 (2018).

## Acknowledgements

L.Z. and L.W. are supported by the National Natural Science Foundation of China (Grant No. 62036002). J.D. is supported by the National Natural Science Foundation of China (Grant No. 61703082) and the Fundamental Research Funds for the Central Universities (Grant No. N2004004).

## Author contributions

L.Z., B.W., and L.W. conceived and designed the project. All authors performed the research. L.Z., B.W., and L.W. performed theoretical calculations. L.Z. performed numerical calculations. L.Z. and L.W. wrote the manuscript. B.W. and J.D. edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23548-4>.

**Correspondence** and requests for materials should be addressed to L.W.

**Peer review information** *Nature Communications* thanks Alberto Antonioni and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021