Reinforcement learning account of network reciprocity

Takahiro Ezaki[1], Naoki Masuda[2]*

1 PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama, Japan,
2 Department of Engineering Mathematics, University of Bristol, Clifton, Bristol, United Kingdom

* naoki.masuda@bristol.ac.uk

Abstract

Evolutionary game theory predicts that cooperation in social dilemma games is promoted when agents are connected as a network. However, when networks are fixed over time, humans do not necessarily show enhanced mutual cooperation. Here we show that reinforcement learning (specifically, the so-called Bush-Mosteller model) approximately explains the experimentally observed network reciprocity and the lack thereof in a parameter region spanned by the benefit-to-cost ratio and the node's degree. Thus, we significantly extend previously obtained numerical results.

Introduction

Human society is built upon cooperation among individuals. However, our society is full of social dilemmas, where cooperative actions, which are costly to individuals, appear to be superseded by non-cooperative, selfish actions that exploit cooperative others [1–3]. There are several mechanisms that explain cooperative behavior in social dilemma situations [4–6]. The evolutionary game theory has provided firm evidence that static networks enhance cooperation as compared to well-mixed populations under generous conditions, with the effect being called spatial reciprocity (in the case of finite-dimensional networks) and network reciprocity (in the case of general networks) [4, 7–10]. This finding is in alignment with broadly made observations that humans as well as animals interact on contact networks where a node is an individual [11–13].

However, a series of laboratory experiments using human participants involved in the prisoner's dilemma game (PDG) has produced results that are not necessarily consistent with spatial and network reciprocity. In fact, the structure of networks (e.g., scale-free, random, and lattice) did not correlate with the propensity of human cooperation in the PDG [14–20]. In contrast, Rand *et al.* have shown that humans present network reciprocity if the benefit-to-cost ratio, a main parameter of the PDG, is larger than the degree of nodes in the network (i.e., number of neighbors per player) [21], which is consistent with the prediction of evolutionary game theory [8]. Note that the earlier experimental studies used smaller benefit-to-cost ratio values [14–20].

The theoretical results in Ref. [8] are derived from the probability of fixation of cooperation, i.e., the probability that a unanimity of cooperation is reached before that of defection under weak selection (i.e., the difference between the strength of cooperator and that of defector is

assumed to be small). While theoretically elegant, unanimity under weak selection may be different from the population dynamics taking place in laboratory experiments with human participants, such as those in Ref. [21]. (However, see [22] for conditions for cooperation that are derived in the case of infinite populations and therefore no fixation, assuming replicator dynamics.) In laboratory experiments, the unanimity of cooperators is hard to be reached. The aim of the present study is to look for an alternative mechanism that explains behavioral results under the PDG on networks.

We hypothesize that a type of reinforcement learning implemented as a strategy of players produces game dynamics that are consistent with the aforementioned experimental results regarding network reciprocity. In particular, aspiration-based reinforcement learning [23–27], with which players modulate their behavior based on the magnitude of the earned reward relative to a threshold, has been successful in explaining conditional cooperation behavior and its variants called moody conditional cooperation [28, 29]. Furthermore, aspiration-based reinforcement learning, not evolutionary game theory, yielded the absence of network reciprocity in numerical simulations [30]. This result is consistent with those showing that outcomes of aspiration-based learning and those of evolutionary dynamics are intrinsically different [31, 32]. In the present paper, we vary the benefit-to-cost ratio and the node's degree, two key parameters in the discussion of network reciprocity in the literature, to show that aspiration-based reinforcement learning gives rise to network reciprocity under the conditions consistent with the previous experimental study [21]. In this way, we significantly extend the previous numerical results [30].

Model

Prisoners' dilemma game on networks

Consider players placed on nodes of a network. They repeatedly play the donation game, which is a special case of the PDG, over $t_{\max}$ rounds as follows. In each round, each player selects either to cooperate (C) or defect (D), and a donation game occurs on each edge in both directions. The submitted action (i.e., C or D) is consistently used toward all neighbors. On each edge, a cooperating player pays cost c to benefit the other player by b. If a player does not cooperate (i.e., D), both the focal player and the other player get nothing. We impose $b > c > 0$. For example, if both players constituting an edge cooperate, both gain $b - c$. Each player is assumed to have k neighbors. Therefore, a player submitting C loses $-kc$ and gains b multiplied by the number of cooperating neighbors. After the donation game has taken place bidirectionally on all edges, each player's final payoff in this round is determined as the payoff that the focal player has gained, averaged over the k neighbors.

Static- and shuffled-network treatments

We compare the propensity of cooperation between static and dynamically shuffled networks, mimicking the situation of a laboratory experiment [21]. In both static- and shuffled- network treatments, the network in each round is a ring network in which each node has k neighbors, where k is an even number (Fig 1). Each player is adjacent to $k/2$ players on each side on the ring. In the static-network treatment, the position of the players is fixed throughout all the rounds. In the shuffled-network treatment, while the network structure is fixed over rounds, we randomize the position of all the players after each round.

BM model

We consider players that obey the Bush-Mosteller (BM) model of reinforcement learning to update actions over rounds [23–25, 27]. We use the following variant of the BM model [29,
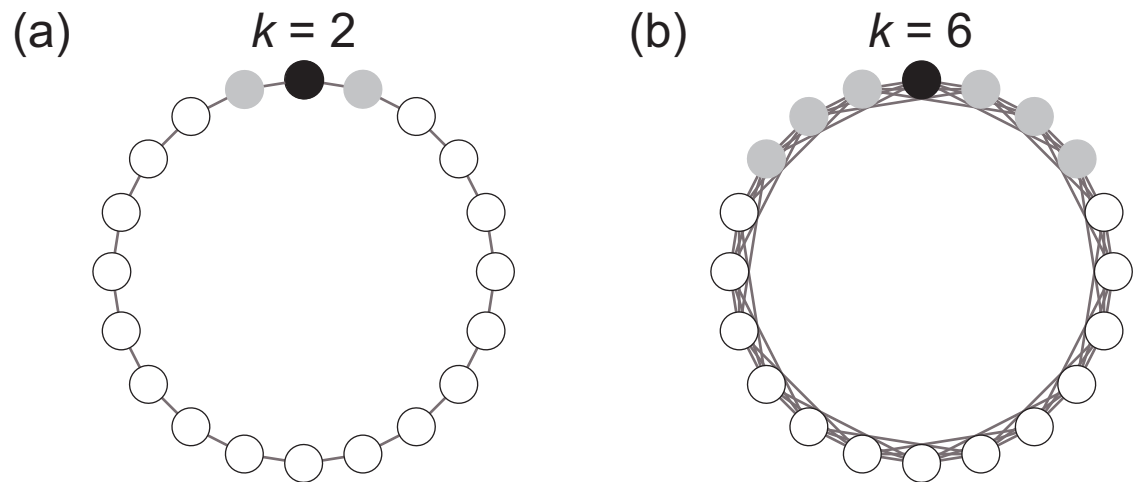
Fig 1. Ring networks composed of N = 20 players. The player represented by a black circle is adjacent to k players represented by gray circles. (a) k = 2. (b) k = 6.

33]. Each player has the intended probability of cooperation, p_t ($t = 1, \ldots, t_{\max}$) as the sole internal state. Probability p_t is updated in response to the payoff obtained in the previous round, denoted by r_{t-1}, and the previous action, denoted by a_{t-1}, as follows:

$$
p_t = \begin{cases}
p_{t-1} + (1 - p_{t-1})s_{t-1} & (a_{t-1} = \mathrm{C}, s_{t-1} \geq 0), \\
p_{t-1} + p_{t-1}s_{t-1} & (a_{t-1} = \mathrm{C}, s_{t-1} < 0), \\
p_{t-1} - p_{t-1}s_{t-1} & (a_{t-1} = \mathrm{D}, s_{t-1} \geq 0), \\
p_{t-1} - (1 - p_{t-1})s_{t-1} & (a_{t-1} = \mathrm{D}, s_{t-1} < 0).
\end{cases}
\tag{1}
$$

In Eq (1) the stimulus, denoted by $s_{t-1} \in (-1, 1)$, is defined by

$$
s_{t-1} = \tanh\left[\beta(r_{t-1} - A)\right],
\tag{2}
$$

where $\beta > 0$ and A are the sensitivity parameter and aspiration level, respectively. The action selected in the previous round is reinforced if the realized payoff is larger than the aspiration level, i.e., $r_{t-1} - A > 0$. Conversely, if the payoff is smaller than the aspiration level, the previous action is suppressed. For example, when a player submitted C in the previous round and the obtained payoff was larger than the aspiration level, the stimulus is positive. Then, the probability of cooperation is increased in the next round [according to the first line in the RHS of Eq (1)]. Note that the updating scheme [Eq (1)] guarantees $p_t \in (0, 1)$ if $p_1 \in (0, 1)$. We set $p_1 = 0.8$, which roughly agrees with the observations made in the previous laboratory experiments [14, 17, 21].

In each round, players are assumed to misimplement the action to submit the action opposite to the intention (i.e., D if the player intends C, and C if the player intends D) with probability ϵ [29, 33–35]. Thus, the actual probability of cooperation is given by $\tilde{p}_t = p_t(1 - \epsilon) + (1 - p_t)\epsilon$. In this way, even defectors that are satisfied with their D action sometimes cooperate. This behavior is not produced by the variation on β.
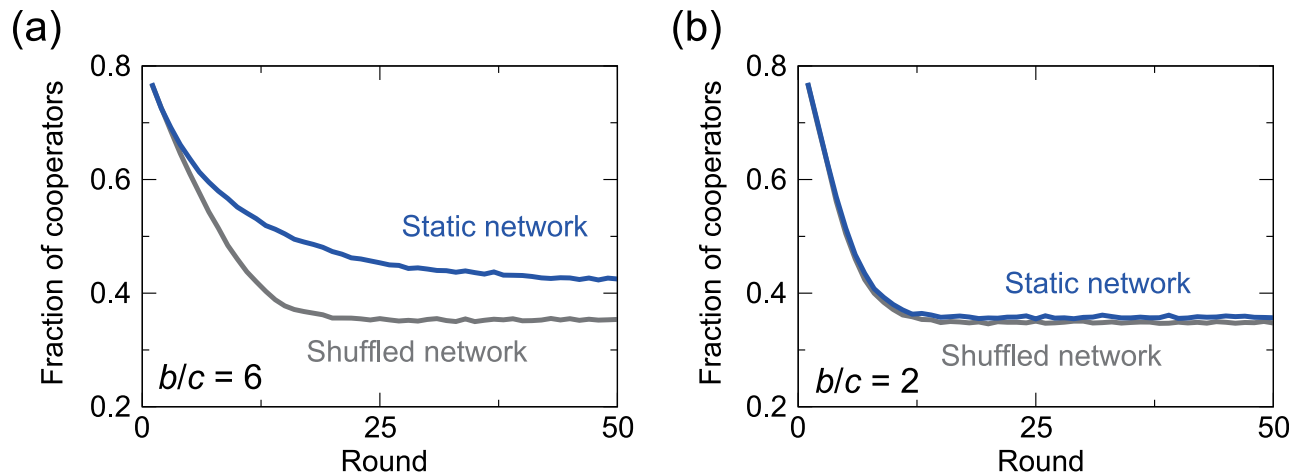
Fig 2. Fraction of cooperative players in each round, averaged over 10^3 simulations. We set $k = 2$, $N = 100$, $t_{max} = 50$, $\beta = 0.2$, $A = 1.0$, and $\epsilon = 0.05$. (a) $b/c = 6$. (b) $b/c = 2$.

Results

We consider two values of b/c, i.e., $b/c = 2$ and 6 by setting $(b, c) = (2, 1)$ and $(b, c) = (6, 1)$, respectively.

Numerically calculated fractions of cooperative players are compared between the two treatments in Fig 2. When the node's degree, k, is small (i.e., $k = 2$) and b/c is large (i.e., $b/c = 6$), cooperation is more frequent in the static-network treatment than the shuffled-network treatment. This result is consistent with the previous experimental results [21]. When $b/c = 2$, this effect is not observed, which is also consistent with the experimental results [14–21].

To examine the robustness of the results shown in Fig 2, we carried out simulations for a region of the $A−\epsilon$ parameter space and four values of k. We did not vary β (= 0.2) because β did not considerably alter the behavior of the players unless it took extreme values [29]. With $b/c = 2$, the fraction of cooperative players averaged over the first 25 rounds is shown in Fig 3(a) and 3(b) for the static and shuffled networks, respectively. The difference between the two types of networks, shown in Fig 3(c), is small in the entire parameter region, in particular for large k, suggesting a marginal effect of network reciprocity. In contrast, when $b/c = 6$, the fraction of cooperators is larger in the static-network than the shuffled-network treatment in a relatively large region of the $A−\epsilon$ parameter space [Fig 3(e), 3(f) and 3(g)]. As k increases, the difference between the two treatments decreases. In summary, a static as opposed to shuffled network promotes cooperation only when b/c is large and k is small. These results are consistent with the experimental findings [21].

Network reciprocity is attributed to assortative connectivity between cooperative players [7–10]. In other words, cooperation can thrive if a cooperator tends to find other cooperators at the neighboring nodes. To measure this effect, we defined the assortment by $P(C|C; t) − P(C|D; t)$, where $P(C|C; t)$ is the probability that a neighbor of a cooperative player is cooperative in round t, and $P(C|D; t)$ is the probability that a neighbor of a defective player is cooperative in round t [21, 36]. For various values of A and ϵ, the assortment values in the static-network treatment averaged over the first 25 rounds are shown in Fig 3(d) and 3(h) for $b/c = 2$ and $b/c = 6$, respectively. The figures indicate that the assortment tends to be positive when cooperation is more abundant in the static-network than shuffled-network treatment regardless of the value of b/c, suggesting that cooperative players are clustered in these parameter regions. In the shuffled treatment, we confirmed that the assortment was ≈ 0 in the entire parameter region.
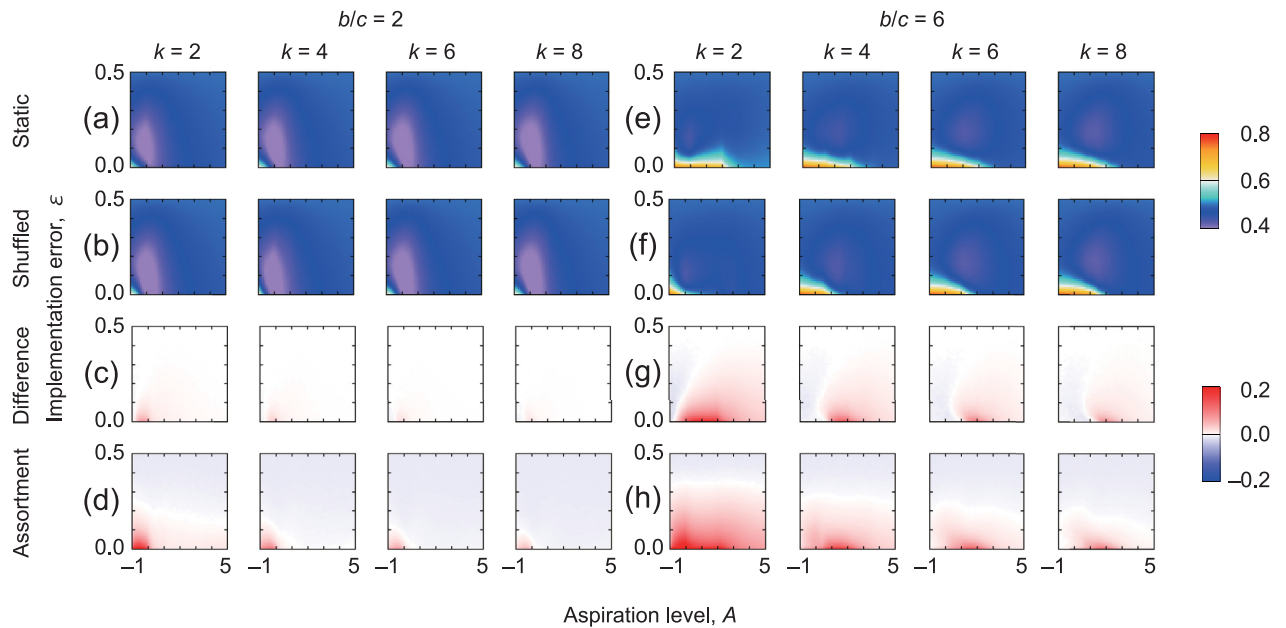
Fig 3. Fraction of cooperative players under the static-network treatment [(a) and (e)] and the shuffled-network treatment [(b) and (f)]. The difference between the fraction of cooperation in the static and shuffled networks is shown in (c) and (g). The assortment for the static networks is shown in (d) and (h). We set $N = 100$ and $\beta = 0.2$. (a)–(d) $b/c = 2$. (e)–(h) $b/c = 6$. To calculate the fraction of cooperators and the assortment, we take averages over the first 25 rounds and 10^3 simulations.

https://doi.org/10.1371/journal.pone.0189220.g003

Conclusions

We have numerically shown that an aspiration-based reinforcement learning model, the BM model, produces network reciprocity if and only if the benefit-to-cost ratio in the donation game is large relative to the node's degree. The results are consistent with the previous experimental findings [14–21]. In addition to network reciprocity, the BM model also accounts for the conditional cooperation, which is hard to explain by evolutionary game theory [28, 29, 37]. Aspiration-based reinforcement learning may be able to describe cooperative behavior of humans and animals in broader contexts. Finally, we remark that, although network reciprocity is not observed in the shuffled-network treatment, in both theory and experiments, dynamic linking treatments that allow players to strategically sever and create links promote cooperation in laboratory experiments [17, 38–41]. Evolutionary game theory predicts cooperation under dynamic linking [42–48]. Reinforcement learning may also account for enhanced cooperation under dynamic linking.

Acknowledgments

Author Contributions

Conceptualization: Naoki Masuda.

Formal analysis: Takahiro Ezaki.

Investigation: Takahiro Ezaki, Naoki Masuda.

Visualization: Takahiro Ezaki.

Writing – original draft: Takahiro Ezaki, Naoki Masuda.

References

1. Dawes RM. Social dilemmas. Annu Rev Psychol. 1980; 31(1):169–193. https://doi.org/10.1146/annurev.ps.31.020180.001125

2. Axelrod R. The Evolution of Cooperation. New York: Basic Books; 1984.

3. Kollock P. Social dilemmas: The anatomy of cooperation. Annu Rev Sociol. 1998; 24(1):183–214. https://doi.org/10.1146/annurev.soc.24.1.183

4. Nowak MA. Five rules for the evolution of cooperation. Science. 2006; 314(5805):1560–1563. https://doi.org/10.1126/science.1133755 PMID: 17158317

5. Sigmund K. The Calculus of Selfishness. Princeton: Princeton University Press; 2010.

6. Rand DG, Nowak MA. Human cooperation. Trends Cogn Sci. 2013; 17(8):413–425. https://doi.org/10.1016/j.tics.2013.06.003 PMID: 23856025

7. Nowak MA, May RM. Evolutionary games and spatial chaos. Nature. 1992; 359(6398):826–829. https://doi.org/10.1038/359826a0

8. Ohtsuki H, Hauert C, Lieberman E, Nowak MA. A simple rule for the evolution of cooperation on graphs and social networks. Nature. 2006; 441(7092):502–505. https://doi.org/10.1038/nature04605 PMID: 16724065

9. Szabó G, Fáth G. Evolutionary games on graphs. Phys Rep. 2007; 446(4–6):97–216. https://doi.org/10.1016/j.physrep.2007.04.004

10. Perc M, Gómez-Gardeñes J, Szolnoki A, Floría LM, Moreno Y. Evolutionary dynamics of group interactions on structured populations: A review. J R Soc Interface. 2013; 10(80):20120997. https://doi.org/10.1098/rsif.2012.0997 PMID: 23303223

11. Easley D, Kleinberg J. Networks, Crowds, and Markets: Reasoning about a Highly Connected World. New York: Cambridge University Press; 2010.

12. Newman M. Networks: An Introduction. New York: Oxford University Press; 2010.

13. Barabási AL. Network Science. Cambridge: Cambridge University Press; 2016.

14. Traulsen A, Semmann D, Sommerfeld RD, Krambeck HJ, Milinski M. Human strategy updating in evolutionary games. Proc Natl Acad Sci USA. 2010; 107(7):2962–2966. https://doi.org/10.1073/pnas.0912515107 PMID: 20142470

15. Cassar A. Coordination and cooperation in local, random and small world networks: Experimental evidence. Games Econ Behav. 2007; 58(2):209–230. https://doi.org/10.1016/j.geb.2006.03.008

16. Grujić J, Fosco C, Araujo L, Cuesta JA, Sánchez A. Social experiments in the mesoscale: humans playing a spatial prisoner's dilemma. PLOS ONE. 2010; 5(11):e13749. https://doi.org/10.1371/journal.pone.0013749 PMID: 21103058

17. Rand DG, Arbesman S, Christakis NA. Dynamic social networks promote cooperation in experiments with humans. Proc Natl Acad Sci USA. 2011; 108(48):19193–19198. https://doi.org/10.1073/pnas.1108243108 PMID: 22084103

18. Suri S, Watts DJ. Cooperation and contagion in web-based, networked public goods experiments. PLOS ONE. 2011; 6(3):e16836. https://doi.org/10.1371/journal.pone.0016836 PMID: 21412431

19. Gracia-Lázaro C, Ferrer A, Ruiz G, Tarancón A, Cuesta JA, Sánchez A, et al. Heterogeneous networks do not promote cooperation when humans play a Prisoner's Dilemma. Proc Natl Acad Sci USA. 2012; 109(32):12922–12926. https://doi.org/10.1073/pnas.1206681109 PMID: 22773811

20. Grujić J, Röhl T, Semmann D, Milinski M, Traulsen A. Consistent strategy updating in spatial and non-spatial behavioral experiments does not promote cooperation in social networks. PLOS ONE. 2012; 7(11):e47718. https://doi.org/10.1371/journal.pone.0047718 PMID: 23185242

21. Rand DG, Nowak MA, Fowler JH, Christakis NA. Static network structure can stabilize human cooperation. Proc Natl Acad Sci USA. 2014; 111(48):17093–17098. https://doi.org/10.1073/pnas.1400406111 PMID: 25404308

22. Ohtsuki H, Nowak MA. The replicator equation on graphs. J Theor Biol. 2006; 243(1):86–97. https://doi.org/10.1016/j.jtbi.2006.06.004 PMID: 16860343

23. Bush RR, Mosteller F. Stochastic Models for Learning. New York: John Wiley & Sons, Inc.; 1955.

24. Rapoport A, Chammah AM. Prisoner's Dilemma: A Study in Conflict and Cooperation. Ann Arbor: The University of Michigan press; 1965.

25. Macy MW. Learning to cooperate: Stochastic and tacit collusion in social exchange. Am J Sociol. 1991; 97(3):808–843. https://doi.org/10.1086/229821

26. Bendor J, Mookherjee D, Ray D. Aspiration-based reinforcement learning in repeated interaction games: An overview. Int Game Theory Rev. 2001; 3:159–174. https://doi.org/10.1142/S0219198901000348

27. Macy MW, Flache A. Learning dynamics in social dilemmas. Proc Natl Acad Sci USA. 2002; 99 (3):7229–7236. https://doi.org/10.1073/pnas.092080099 PMID: 12011402

28. Cimini G, Sánchez A. Learning dynamics explains human behaviour in Prisoner's Dilemma on networks. J R Soc Interface. 2014; 11(94):20131186. https://doi.org/10.1098/rsif.2013.1186 PMID: 24554577

29. Ezaki T, Horita Y, Takezawa M, Masuda N. Reinforcement learning explains conditional cooperation and its moody cousin. PLOS Comput Biol. 2016; 12(7):e1005034. https://doi.org/10.1371/journal.pcbi.1005034 PMID: 27438888

30. Cimini G, Sánchez A. How evolutionary dynamics affects network reciprocity in Prisoner's Dilemma. J Artif Soc Soc Simul. 2015; 18(2):22. https://doi.org/10.18564/jasss.2726

31. Du J, Wu B, Altrock PM, Wang L. Aspiration dynamics of multi-player games in finite populations. J R Soc Interface. 2014; 11(94):20140077–20140077. https://doi.org/10.1098/rsif.2014.0077 PMID: 24598208

32. Du J, Wu B, Wang L. Aspiration dynamics in structured population acts as if in a well-mixed one. Sci Rep. 2015; 5(1):8014. https://doi.org/10.1038/srep08014 PMID: 25619664

33. Masuda N, Nakamura M. Numerical analysis of a reinforcement learning model with the dynamic aspiration level in the iterated prisoner's dilemma. J Theor Biol. 2011; 278(1):55–62. https://doi.org/10.1016/j.jtbi.2011.03.005 PMID: 21397610

34. Nowak M, Sigmund K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature. 1993; 364(6432):56–58. https://doi.org/10.1038/364056a0 PMID: 8316296

35. Nowak MA, Sigmund K, El-Sedy E. Automata, repeated games and noise. J Math Biol. 1995; 33 (7):703–722. https://doi.org/10.1007/BF00184645

36. van Veelen M. Group selection, kin selection, altruism and cooperation: When inclusive fitness is right and when it can be wrong. J Theor Biol. 2009; 259(3):589–600. https://doi.org/10.1016/j.jtbi.2009.04.019 PMID: 19410582

37. Horita Y, Takezawa M, Inukai K, Kita T, Masuda N. Reinforcement learning accounts for moody conditional cooperation behavior: experimental results. Sci Rep. 2017; 7:39275. https://doi.org/10.1038/srep39275 PMID: 28071646

38. Fehl K, van der Post DJ, Semmann D. Co-evolution of behaviour and social network structure promotes human cooperation. Ecol Lett. 2011; 14(6):546–551. https://doi.org/10.1111/j.1461-0248.2011.01615.x PMID: 21463459

39. Wang J, Suri S, Watts DJ. Cooperation and assortativity with dynamic partner updating. Proc Natl Acad Sci USA. 2012; 109(36):14363–14368. https://doi.org/10.1073/pnas.1120867109 PMID: 22904193

40. Jordan JJ, Rand DG, Arbesman S, Fowler JH, Christakis NA. Contagion of cooperation in static and fluid social networks. PLOS ONE. 2013; 8(6):e66199. https://doi.org/10.1371/journal.pone.0066199 PMID: 23840422

41. Shirado H, Fu F, Fowler JH, Christakis NA. Quality versus quantity of social ties in experimental cooperative networks. Nat Commun. 2013; 4:2814. https://doi.org/10.1038/ncomms3814 PMID: 24226079

42. Zimmermann MG, Eguíluz VM, San Miguel M. Coevolution of dynamical states and interactions in dynamic networks. Phys Rev E. 2004; 69(6):065102(R). https://doi.org/10.1103/PhysRevE.69.065102

43. Eguíluz VM, Zimmermann MG, Cela-Conde CJ, San Miguel M. Cooperation and the emergence of role differentiation in the dynamics of social networks. Am J Sociol. 2005; 110(4):977–1008. https://doi.org/10.1086/428716

44. Zimmermann MG, Eguíluz VM. Cooperation, social networks, and the emergence of leadership in a prisoner's dilemma with adaptive local interactions. Phys Rev E. 2005; 72(5):056118. https://doi.org/10.1103/PhysRevE.72.056118

45. Pacheco JM, Traulsen A, Nowak MA. Coevolution of strategy and structure in complex networks with dynamical linking. Phys Rev Lett. 2006; 97(25):258103. https://doi.org/10.1103/PhysRevLett.97.258103 PMID: 17280398

46. Pacheco JM, Traulsen A, Nowak MA. Active linking in evolutionary games. J Theor Biol. 2006; 243 (3):437–443. https://doi.org/10.1016/j.jtbi.2006.06.027 PMID: 16901509

47. Gross T, Blasius B. Adaptive coevolutionary networks: A review. J R Soc Interface. 2008; 5(20):259–271. https://doi.org/10.1098/rsif.2007.1229 PMID: 17971320

48. Perc M, Szolnoki A. Coevolutionary games– A mini review. Biosystems. 2010; 99(2):109–125. https://doi.org/10.1016/j.biosystems.2009.10.003 PMID: 19837129