

Research



Cite this article: Foley M, Forber P, Smead R, Riedl C. 2018 Conflict and convention in dynamic networks. *J. R. Soc. Interface* **15**: 20170835.
<http://dx.doi.org/10.1098/rsif.2017.0835>

Received: 8 November 2017

Accepted: 23 February 2018

Subject Category:

Life Sciences – Mathematics interface

Subject Areas:

computational biology, evolution

Keywords:

game theory, networks, reinforcement learning, evolutionary dynamics, convention

Author for correspondence:

Christoph Riedl
 e-mail: c.riedl@neu.edu

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.4020829>.

Conflict and convention in dynamic networks

Michael Foley¹, Patrick Forber⁵, Rory Smead² and Christoph Riedl^{1,3,4,6}

¹Network Science Institute, ²Department of Philosophy and Religion, ³D'Amore-McKim School of Business, and

⁴College of Computer and Information Science, Northeastern University, Boston, MA, USA

⁵Department of Philosophy, Tufts University, Medford, MA, USA

⁶Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA

id PF, 0000-0002-5060-2187; RS, 0000-0003-2835-3409; CR, 0000-0002-3807-6364

An important way to resolve games of conflict (snowdrift, hawk–dove, chicken) involves adopting a convention: a correlated equilibrium that avoids any conflict between aggressive strategies. Dynamic networks allow individuals to resolve conflict via their network connections rather than changing their strategy. Exploring how behavioural strategies coevolve with social networks reveals new dynamics that can help explain the origins and robustness of conventions. Here, we model the emergence of conventions as correlated equilibria in dynamic networks. Our results show that networks have the tendency to break the symmetry between the two conventional solutions in a strongly biased way. Rather than the correlated equilibrium associated with ownership norms (play aggressive at home, not away), we usually see the opposite host–guest norm (play aggressive away, not at home) evolve on dynamic networks, a phenomenon common to human interaction. We also show that learning to avoid conflict can produce realistic network structures in a way different than preferential attachment models.

1. Introduction

An essential feature of cooperation involves resolving conflicts, and an effective way to resolve conflicts is to adopt a convention that avoids them altogether. Evolutionary game theory provides a way to represent the conflicts associated with the problem of cooperation [1–3], and reinforcement learning can produce adaptive solutions to broad classes of strategic interactions [4–10]. In games of conflict (snowdrift, hawk–dove, chicken) [2] individuals can learn to play a correlated equilibrium, similar to turn-taking, that does better than the mixed Nash equilibrium [11–14]. These solutions, often understood as an important aspect of conventions, play a central role in evolutionary accounts of coordinated action, territoriality and ownership [12,13,15–21]. However, most studies of convention examine learning or evolution with random interactions or static network structures. In reality, learning affects social networks as well as behaviour. Individuals may seek out or avoid others based on past interactions and this can have profound effects on strategy choices, as work on dynamic networks has shown [22–35], an effect also seen in empirical work on human behaviour [36]. For games of conflict, dynamic networks allow individuals to avoid aggressive neighbours by choosing their interaction partners which tends to promote cooperation [37], although some static network structures work against cooperation [38].

Here, we present novel results regarding the emergence of conventions as correlated equilibrium solutions for games of conflict. We use a coevolutionary model where both network ties and behavioural strategies evolve by reinforcement learning. Few previous studies model coevolution of network and strategy evolving with reinforcement learning [19,39,40] but none have explored how dynamic networks influence the evolution of correlated equilibria. We employ computer simulations to search the entire pay-off space for games of conflict and to explore dynamic behaviour over a wide range of additional model parameters, including population size. We find that correlated equilibria often emerge spontaneously as a natural result of reinforcement learning on

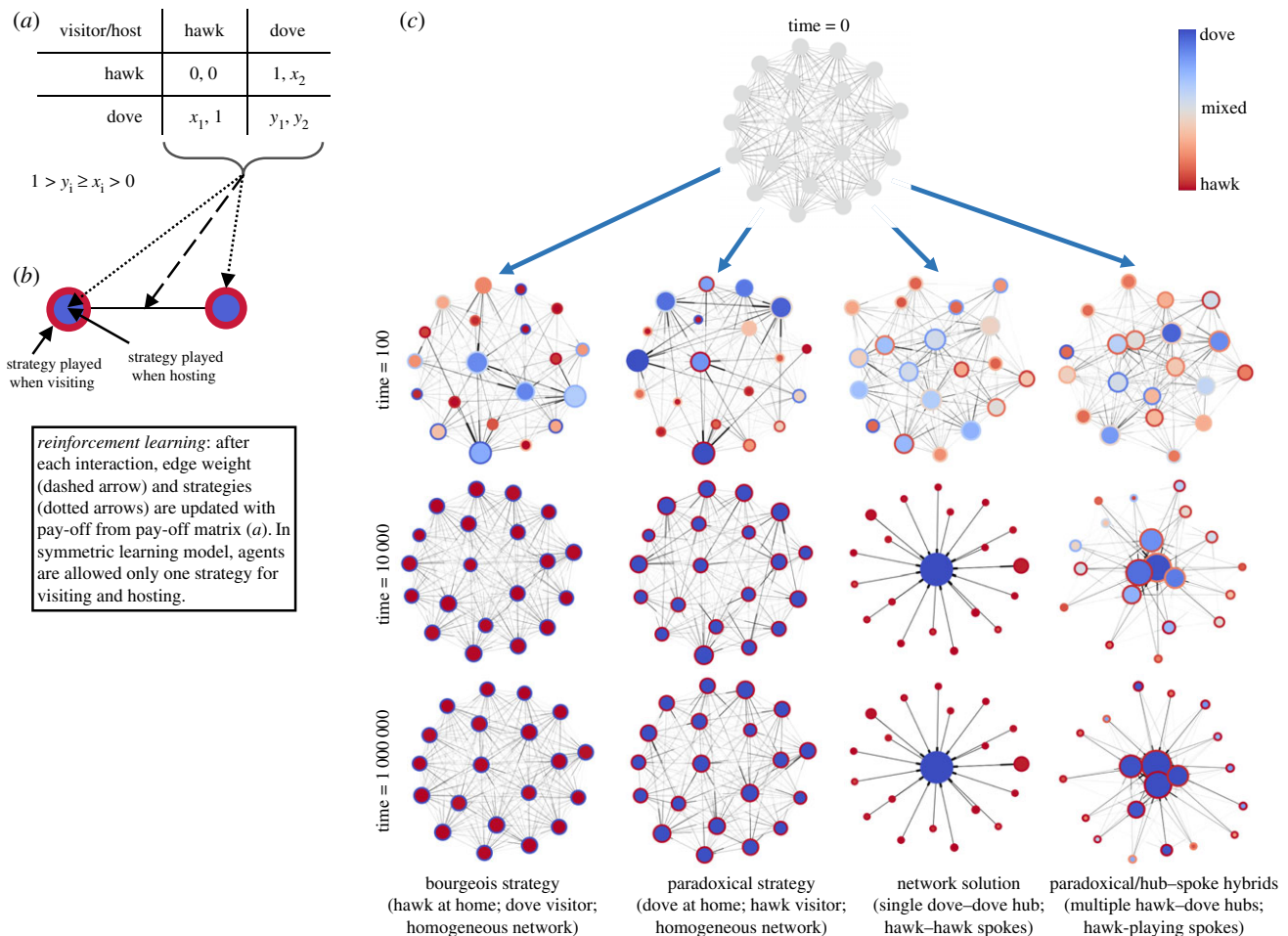


Figure 1. Coevolution of network ties and strategies. (a) Pay-off matrix for the hawk–dove game. (b) Reinforcement learning for host and visitor roles. (c) Starting from uniform homogeneous mixing (fully connected network) and random strategies, populations tend to evolve into one of four stable outcomes depending on early random events and simulation parameters: bourgeois, paradoxical, network or hybrid solutions. Node colours represent strategy weights, node size is proportional to the number of expected visitors and stronger connections are denoted with darker lines.

networks, and that dynamic networks strongly favour the so-called paradoxical solution where individuals act aggressively away but not at home. This result suggests that the prevalence of territoriality and ownership behaviours is not primarily a matter of convention. Our results also have relevance to general mechanisms of network formation. Learning to avoid conflict by evolving network ties can produce realistic network structures and thus provides another potential mechanism for social network formation that differs from preferential attachment [26,37,41].

In games of conflict, each agent can play one of two strategies: an aggressive, competitive strategy (hawk) or a more passive, cooperative strategy (dove). For economy of presentation we will use the hawk–dove terminology, but the results apply equally to any game of conflict meeting the pay-offs defined in figure 1a [2]. The highest pay-off results from playing hawk against dove, the lowest from playing hawk against hawk and all players would prefer their opponent to play dove. Agents playing these games either coordinate on who gets the large pay-off, compete for a resource, or engage in a stand-off. In an unstructured population, common evolutionary dynamics (e.g. replicator dynamics) converge on the mixed Nash equilibrium of the game: host plays dove with probability $x_1/(1 - y_1 + x_1)$ and the visitor plays dove with probability $x_2/(1 - y_2 + x_2)$.

Introducing a dynamic network permits individuals to find solutions different than the mixed Nash (summarized in

figure 1c). If network ties are updated symmetrically using reinforcement learning (i.e. both players update their network ties based on their interaction pay-offs as in [39,40]), there is a new potential solution to games of conflict: agents can learn pure strategies while learning to interact with suitable partners. In these cases, a hub-and-spoke network structure tends to emerge. One agent becomes a hub and learns to play a pure dove strategy, whereas other agents in the population learn to play pure hawk and only visit the dove hub. In effect, agents structure their interaction network to avoid conflict, allowing for a kind of cooperation in games of conflict, as previous work has shown [37].

Another potential new solution, in the form of efficient correlated equilibria, can emerge when individuals are able to coordinate their strategies on an external cue or signal. Correlating behaviour on some external cue (say a coin toss) can produce a convention, such as taking turns playing aggressively (e.g. player one plays hawk when heads, dove otherwise; and player two plays hawk when tails, dove otherwise) [11,13]. Indeed, correlated equilibria can be learned by relatively simple learning rules [42–44], and are also readily learned by humans [45]. In the context of networks, correlated equilibria may arise if individuals learn their host and visitor behaviour independently. This enables them to play one way when hosting and another way when visiting, using their role as host versus visitor as the public cue (figure 1b). At one such correlated equilibrium individuals play what is called

the bourgeois strategy, where they play hawk when hosting and dove when visiting. This strategy has been invoked as an explanation for norms of ownership or territoriality [12,15]. That is, norms of ownership may just be a conventional solution to games of conflict, rather than reflecting the intrinsic value of the resource or territory. However, there is another equally efficient solution that uses the paradoxical strategy, where one plays dove when hosting and hawk when visiting. This alternative solution complicates the evolutionary account of ownership and territoriality. In unstructured populations with symmetric pay-offs both solutions are strongly stable and evolve with equal frequency [13]. This presents a problem of symmetry-breaking: why does it seem that the bourgeois strategy is so much more prominent in nature (see, e.g. [46]) when there is another equally viable option? Previous work has argued that individual learners who are sensitive to relatively small inequalities in value of the territory or fighting ability of the current owner may break the symmetry and strongly favour the bourgeois solution [13]. Here we develop a model to investigate the emergence and stability of these correlated equilibria for games of conflict.

2. The model

In our dynamic network model, agents simultaneously learn how to behave and with whom to interact via Roth–Erev reinforcement. Roth–Erev reinforcement learning provides a simple yet psychologically realistic learning dynamic in which an individual's behaviour evolves iteratively in response to the pay-offs obtained through interaction [6]. This contrasts with other models of evolutionary game theory, such as the replicator dynamics, where 'learning' (evolution) occurs globally across the population as individuals reproduce at rates proportional to the success of their behavioural strategies. In reinforcement learning, individuals have weights associated with each possible behaviour and these weights are updated based on the pay-offs received during their interactions (figure 1*a–b*). Unlike imitation learning, reinforcement learning does not require individuals to track the behaviours or pay-offs of other players in the population, a common feature in contagion models [47]; an agent only tracks the pay-offs associated with their own behaviour and updates their learning weights accordingly. Initial weights determine the initial behaviour of the system. For instance, when initial weights are equal and relatively small, we see effectively random behaviour on a uniform network at the start. At every time step each agent visits another agent to play a game, receives a pay-off and updates learning weights. This update can be symmetric—both host and visitor update their network connection and strategy weights simultaneously—or asymmetric—the hosts only update their strategy choices when hosting, whereas the visitors update both their network connection and strategy choices when visiting. The rationale for this particular asymmetry is that while individuals can control their behavioural strategies when hosting or visiting, and who they visit, they cannot control who decides to visit them. Specifically, we implement this asymmetry by having separate learning weights for hosting and visiting roles, allowing individuals the potential to learn different host and visitor strategies. Also, as hosts have no control over who decides to visit, only visitors update their network connections based on the interaction pay-offs.

The asymmetry in network updating allows for correlated equilibria to emerge and evolve with respect to player roles and therefore provides a way to investigate the conventional origins of norms of ownership or territoriality. Also, in contrast with many previous dynamic network models that allow network ties to form or dissolve discretely [23,26,31,37], our model allows network connections to strengthen or weaken as a function of interaction pay-offs. As ties are rarely discrete in social networks [48–50], modelling network connections with reinforcement learning adds an important element of realism. Furthermore, modelling network ties through reinforcement learning avoids the need to fix the number of network ties which can affect the results [51]. Instead, we allow the number (strength) of ties to evolve endogenously in response to learning.

More precisely, we model a set of N agents engaged in pairwise games of conflict across a number of rounds. Each round, every agent chooses one other agent to visit and engages in an interaction. During each interaction the host and visitor independently choose a behaviour, receive a pay-off, and learn via reinforcement. All players have one interaction as a visitor and up to $N - 1$ interactions as a host.

Each agent i has two vectors (w_H, w_D) and (w_h, w_d) with i 's reinforcement weights for each strategy when hosting or visiting, respectively. Each agent i also has a vector representing their reinforcement weights for choice of which player to visit: $(w_{i1}, w_{i2}, \dots, w_{in})$ where w_{ij} represents the weight related to player i visiting player j . Self-visits are not allowed ($w_{ii} = 0$). Initial weights for both strategy choices are set to 1, while network partner weights are uniformly set to $L/(N - 1)$. We adopted the convention of $L = 19$, so network partner weights for the smallest population size we study ($N = 20$) would start at 1 ($w_{ij} = 1$ for all i, j with $i \neq j$ as the baseline). We kept L constant across different population sizes to ensure that the reinforcement learning has a similar speed relative to total initial weights in larger population sizes. Changing L effectively varies network learning speed; larger L values slow the responsiveness of an agent to interaction pay-offs. (See electronic supplementary material for further details about initial weights and varying learning speeds.)

The model has two modular dynamical components: discounting (δ) and errors (ϵ). Discounting involves reducing past learning weights as more reinforcement occurs, effectively allowing agents to forget older strategies and network connections [6,40]. Errors represent mistakes, mutations or noise where an agent selects a strategy or node in the network at random rather than according to their learning weights. Both of these components are known to impact the stability and long-run behaviour of reinforcement learning [39].

When taking an action, for a given agent in a given context (Host, Visitor, Partner Choice), the probability of choosing option s is proportional to the current relevant weights:

$$\Pr(s) = (1 - \epsilon) \frac{w_s}{\sum_{s'} w_{s'}} + \epsilon \frac{1}{|S|}, \quad (2.1)$$

where ϵ is the error rate, S is the relevant set of available choices (i.e. visiting strategies, hosting strategies or choice of player to visit), $s \in S$ and $s' \in S$.

After each instance of learning, the relevant weights are discounted by a factor δ and increased by the received pay-off (π)

$$w'_s = (1 - \delta)w_s + \pi_s, \quad (2.2)$$

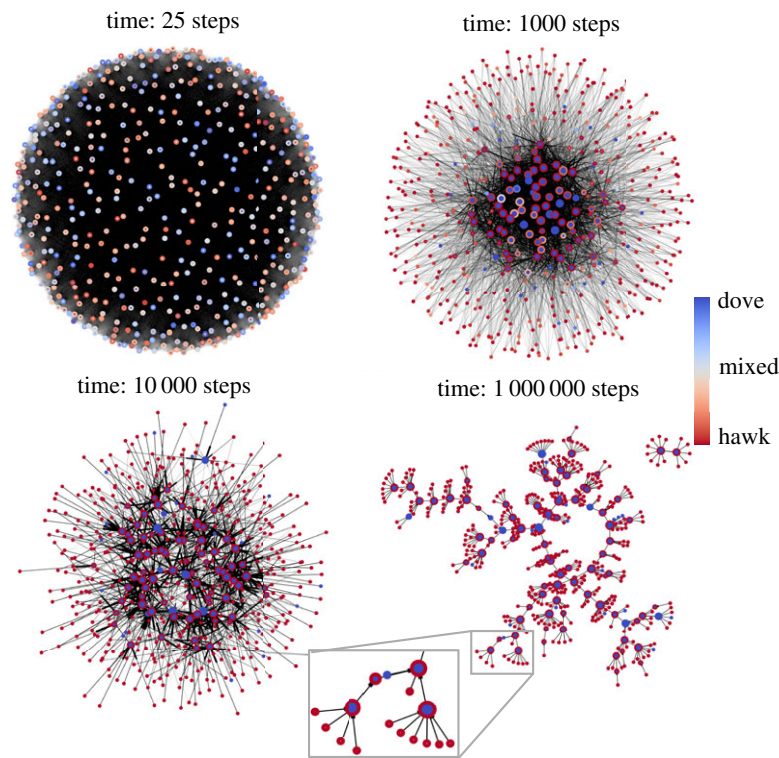


Figure 2. Illustrative example. A representative result of coevolution of network and strategies in a larger population ($N = 500$, discount $\delta = 0.01$, no error, pay-off $x_1 = x_2 = 0.2$, $y_1 = y_2 = 0.6$). Without errors the network approaches a complicated hybrid solution. Node size is proportional to the number of expected visitors.

where w_s' is the weight after updating and π_s is the most recent pay-off related to choosing action s . $\pi_s = 0$ if action s was not chosen. All visits and strategy choices occur simultaneously within a round. All weights are updated simultaneously after each round at a rate of one update per related interaction from that round. Pay-offs received for strategy choice when visiting do not alter the weights for strategy choice when hosting and vice versa. Likewise, pay-offs received for hosting do not affect adjacency weights for choosing which other agents to visit.

3. Results

We explore our model through simulations of reinforcement learning on the asymmetric dynamic network. To provide some context for interpreting the simulation results, let us first describe the sort of solutions we see for games of conflict on dynamic networks. First, there are correlated equilibrium solutions. Note that a population may reach a correlated equilibrium by coordinating their strategies on whether they are host or visitor in one of two ways. Agents may learn to play hawk when hosting and dove when visiting. Hosts always receive the maximal pay-off and visitors always receive x_2 (the bourgeois solution). Alternatively, agents may learn to play dove when hosting and hawk when visiting. This equilibrium grants the visitors the maximal pay-off while the hosts receive x_1 (the paradoxical solution). In these cases, the resulting network is a homogeneous, fully connected graph that approximates a randomly mixing population. Once the correlated equilibrium is established, visiting any agent yields the same pay-off as everyone is deploying the same strategies conditional on the same cues. These correlated equilibrium solutions effectively avoid any conflict between aggressive hawk strategies, and these solutions are commonly found in asymmetric network runs with discounting and errors.

Distinct from correlated equilibrium solutions, there are network solutions that mitigate conflict by using network structure alone: if any agent starts to lean towards dove as host, then agents learn to visit that dove-leaning agent more and also learn to play hawk when they visit. This produces a hub-spoke network structure with one dove host attracting many hawk visitors. In symmetric dynamic network runs we usually see network solutions emerge. Hybrid solutions involve a combination of the paradoxical correlated convention with the network solution where a small number of dove hubs learn the correlated equilibrium and visit each other, while hawk spokes learn a pure (aggressive) strategy and maintain connection to the paradoxical dove-as-host hubs. Hybrid solutions usually emerge for asymmetric network runs with discounting but no errors (e.g. figure 2).

Across all simulations we find that agents arrive at solutions to the games of conflict that fall into one of the four families described above: the two potential conventions (bourgeois and paradoxical), network solutions or hybrid solutions (as in figure 1c). We find no instances of the mixed Nash equilibrium and simulation runs approach one of the four solution families by one million time steps with only very rare exceptions. The relative frequency of outcomes approaching each family varies with the presence of errors, the rate of discounting and the specific values of the pay-offs. Here, we present the results with discounting for both network connections and strategies, as well as with and without errors. These results are qualitatively representative for a constellation of other factors, including larger population sizes, different relative learning speeds for strategy and network connections, and learning with no discounting. Additional results are discussed in the electronic supplementary material.

We conducted an exhaustive search of the entire pay-off space for games of conflict (omitting near boundary cases where y_i or x_i approach 0 or 1). As shown in figure 3a, the paradoxical solution is by far the most prevalent outcome

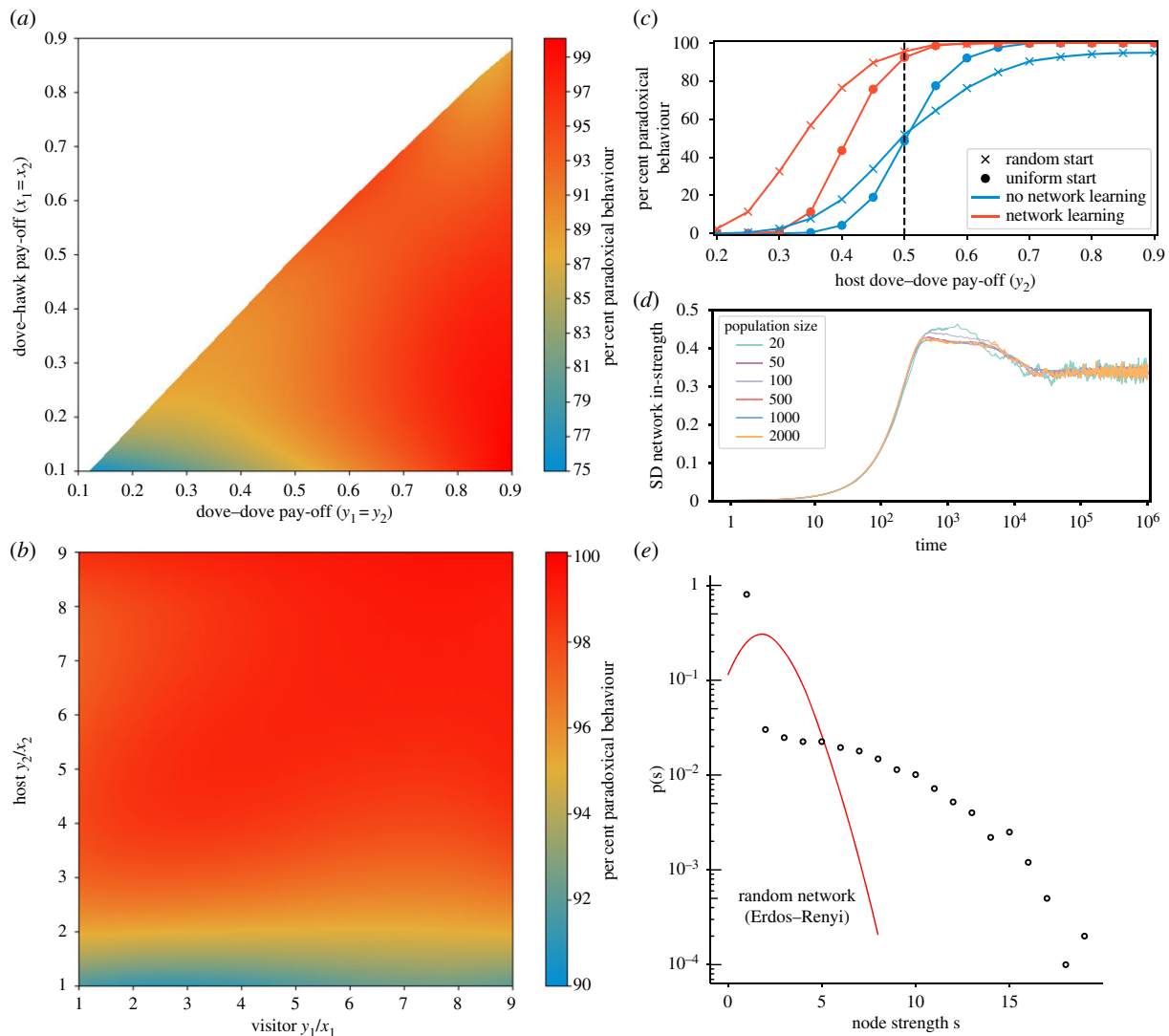


Figure 3. Simulation results. (a) Heatmap representing the proportion of outcomes approaching the paradoxical convention for the space of all possible games of conflict with symmetric pay-offs (discount $\delta = 0.01$, error $\epsilon = 0.01$, $N = 20$). (b) Heatmap representing the proportion of outcomes approaching paradoxical convention for a slice of the asymmetric pay-off space where $y_2 + x_2 = 1$, $y_1 + x_1 = 1$. The x-axis shows the ratio of dove–dove to dove–hawk pay-offs for visitors, while the y-axis shows the same ratio for the hosts' pay-offs. This ratio represents the relative incentive to find a dove partner while playing dove ($\delta = 0.01$, $\epsilon = 0.01$, $N = 20$). (c) The proportion of paradoxical behaviour produced by simulations runs (1000 seeds) for random versus uniform initial learning weights both with and without network learning. Pay-offs are as follows: $y_1 = 0.5$, $x_1 = 0.4$, y_2 varies from 0.2 to 0.9 (along the x-axis), and $x_2 = y_2 - 0.1$. The vertical dashed line represents symmetrical pay-offs. When $y_2 < y_1$, pay-offs are biased towards the bourgeois strategy, and when $y_2 > y_1$, pay-offs are biased towards the paradoxical strategy. (d) Convergence to homogeneous network in the bourgeois and paradoxical solution across different population sizes. Parameters in this plot are as follows: pay-offs $x_2 = 0.3$, $y_2 = 0.5$, $x_1 = 0.6$, and $y_1 = 0.8$, $\delta = 0.01$, $\epsilon = 0.01$. (e) The degree distribution of networks across simulation results shows heterogeneity of network ties across individuals in the population ($\delta = 0.01$, no error, $N = 1000$, pay-off $x_1 = x_2 = 0.2$, $y_1 = y_2 = 0.6$, average of 10 simulations shown).

of simulation runs for symmetric pay-offs. We also investigated the case where hosts and visitors have different pay-off functions, effectively valuing resources differently when host versus visitor. When pay-offs are biased in this way, such that it may be more valuable to play a particular strategy (hawk or dove) at home rather than away (or vice versa), the paradoxical solution is still prevalent, though less so when pay-offs are strongly biased towards the bourgeois solution (electronic supplementary material, figure 3b). It is important to note that biasing the pay-offs in this way makes the bourgeois solution Pareto optimal and, in many contexts, makes it the preferred or expected equilibrium of evolutionary processes [52,53]. The prevalence of the paradoxical solution in cases where the pay-offs favour the bourgeois solution reveals the influence of dynamic networks on breaking the symmetry of conventions.

Initializing simulations with random rather than uniform starting learning weights on initial strategies further increases

the prevalence of paradoxical results on dynamic networks (figure 3c). As a control case, we examined the model without network learning where agents simply choose whom to visit at random every round while still including asymmetric strategy updating. This is equivalent to a randomly mixing population. As one would expect in this case, we see an even split between bourgeois and paradoxical solutions when pay-offs are symmetrical (figure 3c). When pay-offs are biased towards one solution or another, we see a proportional increase in that solution evolving, with random learning weights introducing more noise. Perhaps surprisingly, the dynamical results are insensitive to population size—we see the same qualitative dynamics in populations of sizes from 20 to 2000 (figure 3d).

Without errors discounting leads to behavioural trapping, agents may eventually forget strategy choices and network connections, sometimes producing suboptimal results [40].

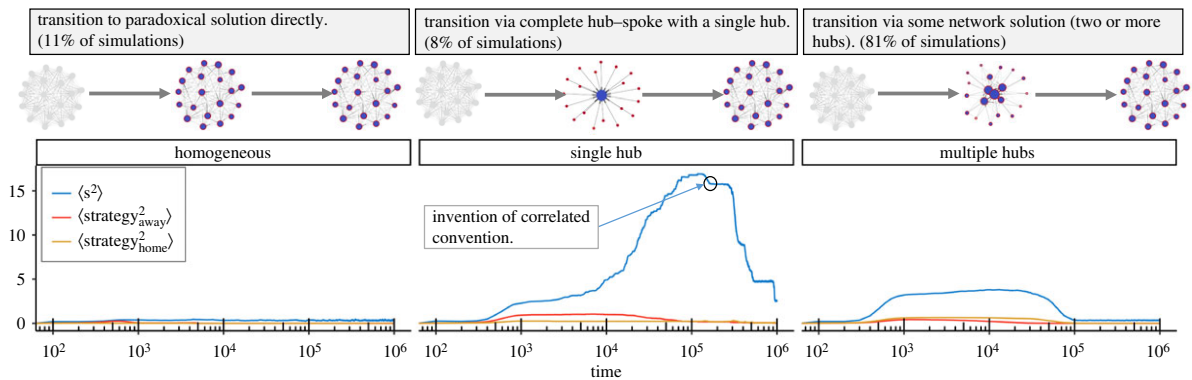


Figure 4. Temporal evolution. Analysis of the possible evolutionary trajectories that produce the paradoxical solution. While some populations find the paradoxical solution right away, most transition to the paradoxical convention by first developing a hub–spoke structure indicative of network or hybrid solutions. Plots show the mean value across all simulations that fall into a specific evolutionary path. Node size shown proportional to the number of expected visitors.

In our model discounting without errors produces stable complex hybrid solutions with only rare exceptions. Figure 2 shows the network evolution and emergence of a hybrid solution for a large population without errors. These networks have heterogeneous degree distributions that are different from random (i.e. Erdos-Renyi) networks (figure 3e). The difference from random networks arises in the tail of the degree distribution, indicating a substantial probability to observe nodes with far more connections than the typical value. This reflects the formation of hubs (either pure dove or paradoxical) that attract many visitors, illustrating a phenomenon found in many real-world networks [41,54,55]. Learning to avoid conflict via reinforcement learning without discounting thus produces a result relevant to current network science: we see realistic network structures produced by an alternative mechanism of social network formation that differs from the preferential attachment and fitness models. Those models are based on a mechanism of growing the network by adding nodes that preferentially attach to already highly connected nodes in the network. Our social network formation model is a self-organizing mechanism that is based on a network with a fixed number of nodes in which tie strengths are modified to avoid conflict.

Errors promote exploration and allow learning to avoid the trapping states. When including an error rate ($\epsilon = 0.001$ or 0.01), the system becomes ergodic, but simulations in the medium-run always show populations close to one of the four solution families. Most often the system is found near one of the conventional solutions with a near homogeneous network. In contrast with figure 2, the inclusion of errors tends to eliminate the dove hubs found in hybrid solutions. Through exploration allowed by errors, the dove hubs tend to find each other and quickly learn the paradoxical solution. Once this starts to catch on, the rest of the population follows, which produces a more homogeneous network playing a correlated equilibrium (figure 4). This result is observed across population sizes (figure 3d).

Of particular interest in reinforcement models is learning speed. Learning speeds can be varied in the model by varying the rate at which weights are accumulated. We systematically investigated the relative speeds of learning for updating network ties and strategies. Previous studies found that varying learning speeds have a noticeable effect on results. Faster network learning speeds tend to promote cooperation [37,39,40]. In our model, we found varying learning speeds had relatively little effect on resulting behaviour. Without errors faster

network learning speed tends to produce hybrid or network solutions, closing off the correlated equilibria. With the addition of errors the network solutions disappear and we see mostly the paradoxical convention. When strategy learning is faster, agents tend to quickly learn one of the conventional solutions and the network tends to stay more homogeneous throughout their evolution (for more detailed discussion of these results see electronic supplementary material).

4. Discussion

The simulation results support a general conclusion: dynamic networks consistently favour the paradoxical solution over the bourgeois solution across a wide range of parameters. The results for probing the possible pay-off space for games of conflict show how frequently simulations produce paradoxical outcomes (figure 3a). Notice the relative rarity of the bourgeois solution, even for cases where host and visitor attach different values to playing hawk as host versus hawk as visitor (figure 3b). Only when pay-offs are very strongly biased in favour of playing dove when visiting, or against dove when hosting, do we find the bourgeois solution more often than the paradoxical solution (figure 3c). If there is no bias or any bias in favour of dove when hosting, or against dove when visiting, the paradoxical solution is far more prominent. A common view on the origin of territoriality or norms of ownership takes the evolution and stability of the bourgeois correlated equilibrium to provide a rationale for aggressively defending territories or resolving conflicts over resources in favour of hosts [12,13,15,18]. In stark contrast, our results show that a host–guest norm, where the host concedes resources to a visitor, evolves far more readily than any norm of territoriality.

The prevalence of the paradoxical solution occurs because the network solution breaks the symmetry between correlated equilibria. In network solutions hubs are strongly reinforced to play dove as host. If a hub agent finds another dove host to visit, then the hub will visit that host more. Once the network connection is strengthened, while hub agents are reinforced to play dove as host by their hawk visitors, they can learn to play hawk as visitor against the other dove host. When this correlated convention is discovered, the rest of the population quickly learns it. Figure 3 shows the possible evolutionary routes to the paradoxical solution. It is only when the bias of the pay-offs towards hawk-when-hosting and dove-when-

visiting is strong enough to overwhelm the influence of the dynamic network that we see the regular evolution of the bourgeois strategy. This is further illustrated by the fact that slower network learning does not lead to the paradoxical solution as frequently as fast network learning (see electronic supplementary material).

This result has an important implication for evolutionary accounts of when or where certain conventions may arise. Note that our model represents scarcity of resources indirectly as variation in pay-off biases (figure 2*d–e*), though we do not consider catastrophic scarcity or interaction outcomes. So, for instance, in the case of territories individuals either are not evicted when conceding to an aggressive opponent, or they can find another territory without severe cost. Also, the model assumes that losses incurred when playing dove against hawks are not overwhelming; individuals do not die or lose all mating opportunities if they concede too many resources. Including scarcity of resources introduces new dynamical elements that may produce different results. If the number of territories is severely limited and a territory is necessary for survival or reproduction, then there is a zero-sum aspect to holding territories that would impact the evolutionary dynamics beyond the game of conflict. The potential significance of these factors would mean that ownership of a territory or control of the resource would confer some intrinsic value, and norms of ownership would not play a merely conventional role in avoiding conflict. Our model reveals that in dynamic networks the prevalence territoriality is not due to relatively small differences in efficiency between the bourgeois and paradoxical conventions. These results support the view that territoriality is prevalent because holding a territory has significant value beyond providing an efficient convention for resolving potential conflicts. If there is no scarcity of resources, or no competition over territories, a host–guest convention is much more likely to emerge even when moderately inefficient. Indeed, this may explain why we tend to see host–guest conventions emerge in many domains of human interaction.

Regarding the evolution of convention more generally, Peyton Young defines a convention as a ‘pattern of behaviour that is customary, expected and self-enforcing’ [16] and other scholars have identified correlated equilibria as central to conventions [17]. Our results demonstrate that correlated equilibria can readily emerge in dynamic networks. In this setting, which agent initiates the interaction provides a ready cue for reinforcement learning and sets the stage for correlated equilibria. Although some accounts of convention involve agents’ expectations or beliefs about the behaviour of other agents [16,17,56], our model shows that learners with no such beliefs or expectations can establish convention-like behaviour (i.e. strategic behaviour uniformly adopted from among equally viable alternatives that is self-enforcing). Furthermore, many potential conventions in game-theoretic contexts pose problems of symmetry since many conventions are equally efficient. The results presented here show that

equally efficient conventions will not evolve with equal probabilities in dynamic networks and thus provides a potential answer to the equilibrium selection problem [57] in these cases.

In conclusion, our results have two important consequences. First, allowing individuals to avoid aggressive neighbours by choosing their interaction partners through a realistic learning mechanism (reinforcement learning) reveals a novel way to resolve games of conflict by promoting hub–spoke structures that, in turn, break the symmetry between correlated equilibria in a strongly biased way. We know that learning in social networks has profound effects on dynamics for other classes of games and most often the results show that dynamic networks favour cooperation in prisoner’s dilemma or stag hunt games [31,37,39,40]. Our study of dynamic networks in games of conflict adds further generality to these results—the network structures that evolve help resolve conflict—and also reveals a surprising tendency to undermine the usual narrative about how the bourgeois convention provides a rationale for norms of ownership or property. Reinforcement learners in dynamic networks have a strong tendency to find the paradoxical solution. Agents only tend to find the bourgeois solution when there are obvious and systematic benefits to territorial behaviour: the interaction pay-offs for games of conflict need to be biased in such a way that the host values playing hawk much more than a visitor. This entails that ownership and territoriality are probably widespread due to the intrinsic importance of holding resources or the value of owning a territory rather than as a convention for avoiding conflict. Second, simple reinforcement learning in a dynamic network can produce realistic network structures in a way that does not presume a growing network with preferential attachment. Exploring how learning to resolve conflict may interact with preferential attachment to influence the structure of human social networks is an exciting prospect for future research.

Data accessibility. Simulations were written in C++ and run for 10^6 rounds of play. Data aggregation and network plots were created using Python and R. When examining asymmetrical pay-offs, two approaches were used. First, we consider cases where $x_i + y_i = 1$ but varied the x_i/y_i ratio between players 1 and 2. Second, we relaxed the requirement of $x_i + y_i = 1$ and considered points varied at regular pay-off intervals. Replication code available on GitHub at <https://github.com/riedlc/ConflictAndConvention>.

Authors’ contributions. M.F., P.F., R.S. and C.R. conceived the experiments. M.F. implemented simulations in C++, M.F. and C.R. conducted the experiment(s), analysed results, and prepared the figures. P.F., R.S. and C.R. wrote the manuscript, M.F., P.F., R.S. and C.R. wrote the supplemental information. All authors contributed to all aspects of the project.

Competing interests. The authors declare no competing financial interests.

Funding. This research was supported in part by Army Research Office grant no. W911NF-14-1-0478, and Office of Naval Research grants N00014-16-1-3005 and N00014-17-1-2542.

References

1. Fudenberg D, Tirole J. 1991 *Game theory*, 1991, vol. 393, p. 12. Cambridge, MA: MIT Press.
2. Weibull JW. 1997 *Evolutionary game theory*. Cambridge, MA: MIT Press.
3. Nowak MA. 2006 Five rules for the evolution of cooperation. *Science* **314**, 1560–1563. (doi:10.1126/science.1133755)
4. Herrnstein RJ. 1970 On the law of effect. *J. Exp. Anal. Behav.* **13**, 243–266. (doi:10.1901/jeab.1970.13-243)
5. Roth AE, Erev I. 1995 Learning in extensive-form games: experimental data and simple dynamic

- models in the intermediate term. *Games. Econ. Behav.* **8**, 164–212. (doi:10.1016/S0899-8256(05)80020-X)
6. Erev I, Roth AE. 1998 Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* **88**, 848–881.
 7. Fudenberg D, Levine DK. 1998 *The theory of learning in games*, vol. 2. Cambridge, MA: MIT Press.
 8. Macy MW, Flache A. 2002 Learning dynamics in social dilemmas. *Proc. Natl Acad. Sci. USA* **99**, 7229–7236. (doi:10.1073/pnas.092080099)
 9. Beggs AW. 2005 On the convergence of reinforcement learning. *J. Econ. Theory*. **122**, 1–36. (doi:10.1016/j.jet.2004.03.008)
 10. Huttegger S, Skyrms B, Tarres P, Wagner E. 2014 Some dynamics of signaling games. *Proc. Natl Acad. Sci. USA* **111**, 10 873–10 880. (doi:10.1073/pnas.1400838111)
 11. Aumann RJ. 1974 Subjectivity and correlation in randomized strategies. *J. Math. Econ.* **1**, 67–96. (doi:10.1016/0304-4068(74)90037-8)
 12. Maynard Smith J. 1982 *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.
 13. Skyrms B. 2014 *Evolution of the social contract*. Cambridge, UK: Cambridge University Press.
 14. Kakade S, Kearns M, Langford J, Ortiz L. 2003 Correlated equilibria in graphical games. In *Proc. of the 4th ACM Conf. on Electronic Commerce*, pp. 42–47. San Diego, California. New York, NY: ACM.
 15. Maynard Smith J, Parker GR. 1976 The logic of asymmetric animal contests. *Nature* **146**, 15–18.
 16. Young HP. 1993 The evolution of conventions. *Econometrica* **61**, 57–84. (doi:10.2307/2951778)
 17. Vanderschraaf P. 1995 Convention as correlated equilibrium. *Erkenntnis* **42**, 65–87. (doi:10.1007/BF01666812)
 18. Kokko H, López-Spulcre A, Morrell LJ. 2006 From hawks and doves to self-consistent games of territorial behavior. *Am. Nat.* **167**, 901–912. (doi:10.1086/504604)
 19. Alexander JM. 2007 *The structural evolution of morality*. Cambridge, UK: Cambridge University Press.
 20. Guala F, Mittone L. 2010 How history and convention create norms: an experimental study. *J. Econ. Psychol.* **31**, 749–756. (doi:10.1016/j.joep.2010.05.009)
 21. Mesterton-Gibbons M, Karabiyik T, Sherratt TN. 2014 The iterated hawk–dove game revisited: the effect of ownership uncertainty on bourgeois as a pure convention. *Dyn. Games Appl.* **4**, 407–431. (doi:10.1007/s13235-014-0111-5)
 22. Jackson MO, Watts A. 2002 On the formation of interaction networks in social coordination games. *Games. Econ. Behav.* **41**, 265–291. (doi:10.1016/S0899-8256(02)00504-3)
 23. Goyal S, Vega-Redondo F. 2005 Network formation and social coordination. *Games. Econ. Behav.* **50**, 178–207. (doi:10.1016/j.geb.2004.01.005)
 24. Santos FC, Pacheco JM. 2005 Scale-free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. Lett.* **95**, 098104. (doi:10.1103/PhysRevLett.95.098104)
 25. Ohtsuki H, Hauert C, Lieberman E, Nowak MA. 2006 A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502–505. (doi:10.1038/nature04605)
 26. Pacheco JM, Traulsen A, Nowak MA. 2006 Coevolution of strategy and structure in complex networks with dynamical linking. *Phys. Rev. Lett.* **97**, 258103. (doi:10.1103/PhysRevLett.97.258103)
 27. Szabó G, Fáth G. 2007 Evolutionary games on graphs. *Phys. Rep.* **446**, 97–216. (doi:10.1016/j.physrep.2007.04.004)
 28. Gross T, Blasius B. 2008 Adaptive coevolutionary networks: a review. *J. R. Soc. Interface* **5**, 259–271. (doi:10.1098/rsif.2007.1229)
 29. Perc M, Szolnoki A. 2010 Coevolutionary games—a mini review. *Biosystems* **99**, 109–125. (doi:10.1016/j.biosystems.2009.10.003)
 30. Wu B, Zhou D, Fu F, Luo Q, Wang L, Traulsen A. 2010 Evolution of cooperation on stochastic dynamical networks. *PLoS ONE* **5**, e11187. (doi:10.1371/journal.pone.0011187)
 31. Rand DG, Arbesman S, Christakis NA. 2011 Dynamic social networks promote cooperation in experiments with humans. *Proc. Natl Acad. Sci. USA* **108**, 19 193–19 198. (doi:10.1073/pnas.1108243108)
 32. Pinheiro FL, Pacheco JM, Santos FC. 2012 From local to global dilemmas in social networks. *PLoS ONE* **7**, e32114. (doi:10.1371/journal.pone.0032114)
 33. Pinheiro FL, Santos FC, Pacheco JM. 2016 Linking individual and collective behavior in adaptive social networks. *Phys. Rev. Lett.* **116**, 128702. (doi:10.1103/PhysRevLett.116.128702)
 34. Allen B, Lippner G, Chen Y-T, Fotouhi B, Momeni N, Yau S-T, Nowak MA. 2017 Evolutionary dynamics on any population structure. *Nature* **544**, 227–230. (doi:10.1038/nature21723)
 35. Wu B, Zhou D, Fu F, Luo Q, Wang L, Traulsen A. 2010 Evolution of cooperation on stochastic dynamical networks. *PLoS ONE* **5**, e11187. (doi:10.1371/journal.pone.0011187)
 36. Fehl K, van der Post DJ, Semmann D. 2011 Co-evolution of behaviour and social network structure promotes human cooperation. *Ecol. Lett.* **6**, 546–551. (doi:10.1111/j.1461-0248.2011.01615.x)
 37. Santos FC, Pacheco JM, Lenaerts T. 2006 Cooperation prevails when individuals adjust their social ties. *PLoS Comput. Biol.* **2**, e140. (doi:10.1371/journal.pcbi.0020140)
 38. Hauert C, Doebeli M. 2004 Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **428**, 643–646. (doi:10.1038/nature02360)
 39. Skyrms B, Pemantle R. 2000 A dynamic model of social network formation. *Proc. Natl Acad. Sci. USA* **97**, 9340–9346. (doi:10.1073/pnas.97.16.9340)
 40. Pemantle R, Skyrms B. 2004 Network formation by reinforcement learning: the long and medium run. *Math. Soc. Sci.* **48**, 315–327. (doi:10.1016/j.mathsocsci.2004.03.007)
 41. Barabási A-L, Albert R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)
 42. Foster DP, Vohara RV. 1997 Calibrated learning and correlated equilibrium. *Games. Econ. Behav.* **21**, 40–55. (doi:10.1006/game.1997.0595)
 43. Hart S, Mas-Colell A. 2000 a simple adaptive procedure leading to correlated equilibrium. *Econometrica* **68**, 1127–1150. (doi:10.1111/1468-0262.00153)
 44. Arifovic J, Boitnott JF, Duffy J. In press. Learning correlated equilibria: an evolutionary approach. *J. Econ. Behav. Organ.* (doi:10.1016/j.jebo.2016.09.011)
 45. Duffy J, Feltovich N. 2010 Correlated equilibria, good and bad: an experimental study. *Int. Econ. Rev. (Philadelphia)* **51**, 701–721. (doi:10.1111/j.1468-2354.2010.00598.x)
 46. Krebs JR, Davies NB. 1993 *An introduction to behavioral ecology*. London, UK: Blackwell.
 47. Dodds PS, Watts DJ. 2005 A generalized model of social and biological contagion. *J. Theor. Biol.* **232**, 587–604. (doi:10.1016/j.jtbi.2004.09.006)
 48. Granovetter MS. 1973 The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380. (doi:10.1086/225469)
 49. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. 2004 The architecture of complex weighted networks. *Proc. Natl Acad. Sci. USA* **101**, 3747–3752. (doi:10.1073/pnas.0400087101)
 50. Yook S-H, Jeong H, Barabási A-L, Tu Y. 2001 Weighted evolving networks. *Phys. Rev. Lett.* **86**, 5835. (doi:10.1103/PhysRevLett.86.5835)
 51. Rand DG, Nowak MA, Fowler JH, Christakis NA. 2014 Static network structure can stabilize human cooperation. *Proc. Natl Acad. Sci. USA* **111**, 17 093–17 098. (doi:10.1073/pnas.1400406111)
 52. Sandholm W. 2010 *Population games and evolutionary dynamics*. Cambridge, MA: MIT Press.
 53. Szabó G, Borsos I. 2016 Evolutionary potential games on lattices. *Phys. Rep.* **624**, 1–60. (doi:10.1016/j.physrep.2016.02.006)
 54. Clauset A, Shalizi CR, Newman ME. 2009 Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703. (doi:10.1137/070710111)
 55. Watts DJ, Dodds PS. 2007 Influentials, networks, and public opinion formation. *J. Consum. Res.* **34**, 441–458. (doi:10.1086/518527)
 56. Lewis D. 1969 *Convention*. Cambridge, MA: Harvard University Press.
 57. Samuelson L. 1997 *Evolutionary games and equilibrium selection*. Cambridge, MA: MIT Press.