Partners and rivals in direct reciprocity

Christian Hilbe[1,2], Krishnendu Chatterjee[2], Martin A. Nowak[1,3]
[1]Program for Evolutionary Dynamics, Harvard University, Cambridge MA 02138, USA
[2]IST Austria, 3400 Klosterneuburg, Austria
[3]Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge MA 02138, USA

Reciprocity is a major factor in human social life and accounts for a large part of cooperation in our communities. Direct reciprocity arises when repeated interactions occur between the same individuals. The framework of iterated games formalizes this phenomenon. Despite being introduced more than five decades ago, the concept keeps offering beautiful surprises. Recent theoretical research driven by new mathematical tools has proposed a remarkable dichotomy among the crucial strategies: successful individuals either act as partners or as rivals. Rivals strive for unilateral advantages by applying selfish or extortionate strategies. Partners aim to share the payoff for mutual cooperation, but are ready to fight back when being exploited. Which of these behaviors evolves, depends on the environment. Whereas small population sizes and a limited number of rounds favor rivalry, partner strategies are selected when populations are large and relationships stable. Only partners allow for evolution of cooperation, while the rivals' attempt to put themselves first leads to defection.

In the 1950s, when Merrill Flood and Melvin Dresher wanted to test novel solution concepts of game theory[1,2], they asked colleagues at the RAND corporation to play several rounds of various two-player games[3]. One of those games had a peculiar quality: by maximizing their own payoffs, players would end up in a situation that is detrimental for both. Since then, the Prisoner's Dilemma[4] (PD) has become a major paradigm to study strategic behavior in social dilemmas. It is presented as a game in which two players, say Alice and Bob, can either cooperate or defect (**Fig. 1a**). If both cooperate, they each get the reward, R, which exceeds the punishment payoff, P, when both defect. But if one player defects while the other cooperates, the defector gets the highest payoff T (temptation), whereas the cooperator ends up with the lowest payoff S (the sucker's payoff). The game is a PD if $T > R > P > S$. No matter what Alice does, Bob maximizes his payoff by defecting. Thus, defection is the only Nash equilibrium.

Pure defection, however, was not the outcome Flood and Dresher observed in their experiment. Instead their participants seemed to become more cooperative over time. When confronted with those results, John Nash argued that the experimental game was not a Prisoner's Dilemma, but a repeated Prisoner's Dilemma[3]. Repeated games allow for reciprocity[5–7]. Players have additional strategic options: they can react to the outcomes of previous rounds, they can reward cooperating co-players by cooperating in the future, they can punish defecting co-players by defecting in the future. Reward and punishment are intrinsic properties of repeated games.

Direct reciprocity is a mechanism for the evolution of cooperation[8], based on the concept that my behavior towards you depends on our previous interactions. To study direct reciprocity, assume that after each round of the PD, there is another one with probability δ. Equivalently, we could assume that there are infinitely many rounds, but future payoffs are discounted by δ. When the game is repeated, the set of feasible strategies is huge. Instead of merely deciding whether to cooperate or to defect in a single interaction, a strategy needs to specify what to do in every round, given the previous history of interactions. To do so, Alice might for example condition her next action on whether Bob has cooperated in the previous round. Alternatively, Alice might cooperate if the running average payoffs of the two players falls within a certain range[9].

Iterated games and the Folk theorem

Repeated interactions allow cooperation to be stable[10]. To see why, assume Alice and Bob can choose between two possible strategies, always defect (ALLD) and Tit-for-Tat (TFT) (**Fig. 1b**). ALLD defects in every round. TFT cooperates in the first round and then does whatever the opponent did in the previous round (**Fig. 2**). When both players use TFT, they get payoff R in every round. If Alice instead switches to ALLD, she obtains $T > R$ in the first round but $P < R$ in all subsequent rounds. If future interactions are sufficiently likely, Alice's short run advantage is not worth her long time loss.

This logic of reciprocity is simple, but it has been effective to understand when individuals cooperate under a 'shadow of the future'. Repeated games have been employed to study topics as diverse as collusion[11,12], venture capitalism[13], arms races[14,15], food-sharing[16,17] and predator inspection[18]. Computer scientists and mathematicians have been interested in the computational complexity of finding best responses[19,20].

Although the rules of the game are simple to describe, the outcome is complex to predict. On one hand, the repeated PD allows for many different equilibria. The folk theorem[21,22] guarantees that any feasible average payoff can arise in equilibrium, provided that players get at least the mutual defection payoff P (**Fig. 1c**). On the other hand, in evolving populations none of those equilibria are evolutionarily stable[23–26]. For example, if all members of a population apply TFT, individuals using 'always cooperate' (ALLC) fare just as well. Thus, ALLC may spread through neutral drift, favoring the subsequent invasion of ALLD[7]. But ALLD is not evolutionarily stable either. It can be neutrally invaded by Suspicious Tit-for-Tat (STFT), which defects in the first round and plays like TFT thereafter. Once STFT is common, more cooperative strategies can take over. Such neutral, stepping-stone invasions are always possible[27,28], unless there is a positive probability of mistakes[29,30]. Moreover, when the dynamics of strategies in a population is modeled as a stochastic process, chance events during mutation and selection may help a mutant strategy to invade even if it is initially at a disadvantage (**Fig. 1d**). Cooperation thus comes and goes in cycles[31–33]. Periods of defection alternate with periods of cooperation, and the respective length of these periods determines how likely we observe cooperation over time[34–39].

Classical strategies for the repeated prisoner's dilemma

In absence of a universally optimal strategy, research has focused on identifying cooperative strategies that perform well in a broad range of scenarios[40–50]. The field owes much of its early momentum to Robert Axelrod , who invited experts to submit programs to play the repeated PD in a computerized round-robin tournament[40]. The shortest program, TFT, submitted by Anatol Rapaport[4] achieved the highest average score, although it did not win any pairwise encounter. Axelrod attributed the success of TFT to four appealing properties: TFT is never is the first to defect, it responds to defection by defecting, it returns to cooperation if the co-player does so, and it is easy for other players to comprehend it. A recent mathematical analysis has shown that the simple imitation rule employed by TFT makes it "unbeatable" in social dilemmas with two actions: against TFT, no opponent can achieve arbitrarily high payoff advantages[51].

But TFT is not as superior as these results suggested. Its success in Axelrod's tournament critically depends on the participating strategies and on the methods used to determine the winner[52]. For example, the strategy Tit-for-two-Tats (TF2T) would have won the first tournament if it only was submitted[40]. TF2T only defects if the co-player has defected in the previous two rounds (**Fig. 2e**). Moreover, a strict retaliator like TFT is unable to correct errors: if players occasionally make mistakes, cooperation between two TFT players breaks down[53–55].

An alternative approach to tournaments is to let evolution decide which strategies prevail[42,43]. Consider a population of players, each one equipped with a specific strategy. Over time, successful strategies spread, either because they reproduce faster or they are imitated more often[56,57]. In addition, mutation or random exploration introduce novel strategies. When modeling such evolutionary processes for the iterated PD, the enormous number of possible strategies makes it often necessary to constrain the available set of strategies. One assumption is that players are reactive: when deciding whether to cooperate in the next round, they only consider the opponent's move in the very last round. Reactive strategies are described a triplet (y, p, q). Here, y is the probability to cooperate in the first round, p is the probability to cooperate if the co-player has cooperated in the previous round and q is the probability to cooperate if the co-player has defected[58]. Although reactive strategies contain both ALLD and TFT, stochastic simulations typically favor a more lenient strategy. Evolutionary trajectories often lead from ALLD to TFT and from there to Generous Tit for Tat (GTFT)[42]. When Alice applies GTFT, she always cooperates in the first round and after rounds in which Bob has cooperated. But when Bob defects, Alice still cooperates with some probability $q > 0$ (**Fig. 2f**). The probability q can be chosen sufficiently large to avoid costly vendettas after an error, but low enough to give ALLD no selective advantage[41].

The evolutionary dynamics change when players additionally take their own previous move into account. Such memory-1 strategies have the form $(p_0;\ p_1, p_2, p_3, p_4)$, where p_0 again represents a player's probability to cooperate in the first round, and the other four numbers are the probabilities to cooperate after the outcomes CC, CD, DC, DD, respectively. The first letter represents the previous action of the focal player, the second letter refers to the action of the co-player. Stochastic memory-one strategies have been extensively used in evolutionary game theory[7]. They are simple enough to be explored with computer simulations[39,48], yet sufficiently complex to encode a variety of interesting behaviors (**Fig. 2**). Once individuals can choose among all memory-1 strategies, evolution often leads to Win-Stay Lose-Shift (WSLS)[43,44]. When Alice applies WSLS, she starts with cooperation; thereafter she repeats her previous action if it yielded at least payoff R in the previous round. If her payoff was less than R, she switches to the opposite action (**Fig. 2g**).

WSLS is the only memory-1 strategy that satisfies three simple principles[48]: It is mutually cooperative, retaliating, and error-correcting. That is, WSLS continues to cooperate after mutual cooperation, it retaliates a co-player's defection by defecting for at least one round, and two WSLS players restore cooperation after at most one round. Due to these principles, WSLS evolves in a wide range of scenarios, provided that mutual cooperation is sufficiently profitable and that the game is iterated for a sufficient number of rounds[43–48].

Zero-determinant strategies

While evolutionary game theory traditionally asks which strategies win in evolving populations, Press & Dyson[59] recently posed a different question: Are there strategies for Alice with which she wins every pairwise encounter with Bob, irrespective of which strategy Bob uses? Moreover, can she achieve this goal in a way that makes it optimal for Bob to cooperate in every round? Surprisingly, the answer to both questions is yes. The argument involves two steps.

First, Press & Dyson described an intriguing subset of memory-1 strategies, the so-called zero-determinant (ZD) strategies. For the derivation of these strategies, a particular matrix plays an important role, which depends on the players' memory-1 strategies. If Alice employs a ZD strategy, the determinant of this matrix becomes zero, which explains the curious name of these strategies. More importantly, Press & Dyson observed that by using a ZD strategy, Alice can enforce a linear relationship between her own and Bob's payoff. The exact shape of this relationship is solely under Alice's control. Second, they showed that among the ZD strategies there are so called 'extortioners'. With an extortionate strategy, Alice can guarantee, for example, that she always gets twice the payoff of Bob, whereas Bob can do no better than cooperating in every single round (see **Box 1**). For that statement to hold, the payoff has to be rescaled such that the payoff for mutual defection is zero.

If Bob is not cooperative from the outset, Alice can employ an extortionate strategy to teach him to cooperate[59,60]. Suppose that Alice is committed to a fixed strategy, while Bob is willing to adapt. Bob may occasionally change his strategy in response to Alice's fixed behavior. When Alice uses an extortionate strategy, any attempt Bob makes to increase his own payoff automatically also increases Alice's payoff. As Bob adapts, he becomes more and more cooperative over time. As a result, both players' payoffs increase, but Alice's payoff increases by twice as much (**Fig. 3a**).

After the discovery of extortionate strategies, several studies have explored their general existence[61–64], their evolutionary performance[65–70], and their relevance for human interactions[71–74]. This work suggests that extortion is feasible in almost any natural setup, even if the social dilemma involves more than two players[61,75], or if players have access to more than two discrete actions[62].

Evolving populations, however, typically do not settle at extortion[65–70]. But extortionate strategies can still act as catalysts for cooperation[65–68]. Since extortioners never lose any direct competition, they can subvert ALLD populations through neutral drift. Once they are common, they quickly give rise to more cooperative strategies. Evolution leads from extortion to generosity[76]. Eventually, successful players provide incentives for mutual cooperation, but they are also willing to accept a *lower* payoff than their opponent when mutual cooperation fails (**Fig. 3b**).

Of Partners and Rivals

Maybe even more important than the discovery of ZD strategies is the new mathematical formalism that comes with them[75–83]. This formalism can be applied more generally to derive relationships between the payoffs players can achieve in repeated games. Using these relationships we find a remarkable dichotomy among the strategies for the iterated PD. Most of the previously discussed strategies fall into one of two classes: they act as rivals or as partners (**Fig. 4**). Partners aim to share the mutual cooperation payoff R with their co-player. Should the co-player not go along, however, they are ready to punish their co-player with lower payoffs. Rivals on the other hand, aim to have a higher payoff than their respective opponent, no matter what the opponent does.

Whether a given strategy qualifies as a partner or rival, depends on the payoff values and the continuation probability. Among reactive and memory-1 strategies for the iterated PD, the sets of partner and rival strategies can be characterized explicitly (**Box 2**). For high continuation probabilities and a considerable benefit to cooperation, the set of partner strategies includes TFT, GTFT, WSLS, and Grim. The set of rival strategies contains ALLD and the class of extortioners.

If the expected number of rounds is finite, subjects cannot be rival and partner at the same time. Partners need to be 'nice' to ensure they yield the mutual cooperation payoff R against like-minded opponents. They are never the first to defect. In contrast, rivals must be 'cautious' to guarantee they cannot be outperformed by any opponent. They are never the first to cooperate. A world of rivals is a world in which everyone defects.

Only in infinitely repeated games without discounting of the future, does TFT offer a compromise between these two classes. In that case, the first round does not matter and TFT is both a partner and a rival: it does not lose out in any pairwise encounter, while still making sure that it yields the mutual cooperation payoff against players of the same kind.

While the two sets of partner and rival strategies comprise many of the well-known strategies, they do not contain all of them. For example, ALLC and TF2T neither qualify as partner or rival. Instead, these strategies could be deemed submissive[83]: players using ALLC or TF2T avoid ever getting a higher payoff than their opponent.

Partners and rivals in evolution

In general, partners and rivals only comprise a small fraction of strategies for the iterated PD. For example, among reactive strategies, partners always need to cooperate if the co-player cooperated in the previous round, while rivals need to defect after a co-player's defection. Due to these constraints, the probability that a randomly chosen reactive strategy is either a partner or a rival is zero (**Fig. 5a**). Nevertheless, evolutionary trajectories visit the vicinity of these two strategy sets disproportionally often (**Fig. 5b–d**). Partner strategies are favored when cooperation yields a high benefit, when populations are sufficiently large, and when errors are rare. However, the reason why partner strategies are favored is different between these cases. High benefits of cooperation are amenable to the evolution of partners because they increase both the set of partner strategies and its basin of attraction[42,58], from which evolution leads toward partner strategies. In contrast, small population sizes leave the set of partner strategies unaffected, but small populations select for spite[34]. When a population contains only a few individuals, successful strategies do not need to yield a high payoff. They only need to guarantee that the own payoff is higher than the payoff of all others. In such cases, rivalry pays.

While the results in **Fig. 5** focus on evolution among reactive strategies in games without discounting, the same conclusions hold for memory-1 strategies with discounting (**Fig. S2**). There we additionally show that rivalry is favored when the game is only played for a few rounds, such that partner strategies cease to exist.

These simulation results can be understood using the concept of evolutionary robustness[76–79]. If a resident population of size N applies a strategy that is evolutionary robust, no mutant strategy can reach fixation with probability higher than the neutral probability $1/N$. In the limit of large populations and no discounting, Stewart & Plotkin have shown that all partner strategies are evolutionary robust, and so is a subset of the rival strategies (called robust self-defectors[77]). The only other robust set of strategies is the set of robust self-alternators, according to which Alice and Bob alternate between cooperation and defection (such that $p_2 = 0$ and $p_3 = 1$). Which behavior will be favored over an evolutionary timescale is surprisingly well predicted by the dynamics among these three strategy sets[77].

Direct reciprocity in the laboratory

Instead of exploring the performance of partners and rivals in virtual populations, one may ask which behaviors human subjects would adopt, using the controlled setting of laboratory experiments. While it is notoriously difficult to infer which strategies subjects apply, based on their revealed actions, experiments provide some evidence for the above evolutionary results. For example, the recent finding that subjects become less cooperative when they focus on the payoffs of their co-players[84] can be interpreted as an illustration of the negative effects of rivalry. Experimental results also seem to be in line with the qualitative trends of **Fig. 5** and **Fig. S2**: Subjects become more cooperative if they can expect to interact in more rounds,[85,86] or when the benefit-to-cost ratio of cooperation is high, and when errors are rare.[87]

Two experiments have aimed to quantify the success of ZD strategies more directly, by matching human participants either with an extortionate or a generous ZD strategy[71,72]. The ZD strategy was implemented by a computer program, but subjects did not obtain any information about the nature of their opponent. While the extortionate program indeed outperformed each human opponent in the direct encounter, it was the generous program that reached on average higher payoffs than the extortioner. For this result, fairness considerations are essential: When being matched with an extortioner, there is a trade-off between gaining high payoffs, which would require the human participants to cooperate, and gaining equal payoffs, which would require them to defect in every round. This trade-off is absent in the generosity treatment: against generous opponents, full cooperation guarantees both, high and equal payoffs. In line with this argument, the concern for fairness vanished when participants were explicitly informed that they are interacting with an abstract computer program, in which case participants were equally cooperative across all treatments[72].

However, extortion may still succeed under appropriate circumstances. A stylized behavioral experiment on climate change negotiations suggests that even if subjects themselves are not extortionate, they may vote for representatives who are[73]. In this way, subjects may reap the benefits of extortion without a need to feel guilty.

Power asymmetries also seem to trigger extortionate behavior: in another experiment, a randomly determined subject was given the option to replace one of her co-players by a currently inactive player every ten rounds of a repeated PD. The replaced player would then become the inactive player, without any opportunity to earn payoffs during that period. Under these rules, subjects with the replacement option learned to take advantage of their superior position. They subtly enforced their opponents' cooperation while being substantially less cooperative themselves[74]. It seems that with great power comes rivalry, instead of responsibility.

Beyond the iterated prisoner's dilemma

While the iterated PD has been the most common model to study direct reciprocity, the repeated games of our daily lives can have slightly different manifestations. Alice and Bob may face different one-shot payoffs[88,89], they may have to make their decisions asynchronously[90–92] or they may have access to richer strategy sets, instead of just having the binary choice between cooperation and defection[93–95]. In other applications, the social dilemma may not only involve Alice and Bob but also Caroline, Dave, and others[37,96]. How do the above results extend to these cases?

None of the results depend on the specific payoff ordering $T > R > P > S$ of the PD. Instead, they readily extend to arbitrary social dilemmas that only satisfy $R > P$, which means mutual cooperation is preferred over mutual defection, and $T > S$ implying that players prefer to be the defector in mixed groups. In that case, the existence of ZD strategies is guaranteed[75], and also the characterization of rival strategies (8) carries over[83]. In particular, these concepts immediately apply to other well-known social dilemmas, such as the snowdrift game (with $T > R > S > P$) and the stag-hunt game (with $R > T > P > S$). Only the characterization of partner strategies (7) requires $T + S < 2R$. Without this condition, alternating cooperation and defection would be the social optimum necessitating a different definition of partner. A numerical analysis for evolutionary games among pure memory-1 strategies supports this view[38]: in social dilemmas, rival and partner strategies are predominant if mutual cooperation is optimal, whereas alternating strategies succeed when $T + S > 2R$.

Similarly, the results also extend to social dilemmas with two actions but multiple players. For example, most of the strategies in **Fig. 2** have direct analogues in the multiplayer case, including TFT[75], WSLS[47,48], and extortioners[61]. Also, the definitions of partner and rival strategies can be extended appropriately. However, in arbitrary multiplayer games the partners among the memory-1 players have only been characterized among deterministic and ZD strategies[75]. For public goods game among memory-k players, Stewart and Plotkin show that a strategy only needs to resist four "extremal" mutant strategies to qualify as a partner[78]. Their analysis also reveals that the relative size of the set of partner strategies increases with the player's memory, but decreases with group size. In line with these analytical results, their simulations confirm that small groups and long memories promote cooperation, and that players learn to expand their memory capacity when given the option[78]. An analogous characterization of rival strategies in multiplayer games is still pending, although such an extension seems feasible using the methods sketched herein (see **Box 2**).

Finally, there has also been substantial progress on social dilemmas with two players but multiple actions. ZD strategies can be characterized for continuous action sets[62]. Moreover, it has been shown that full cooperation can often be stabilized with partner strategies that only make use of two of the n possible actions[79]. At the same time, however, simulations suggest that evolution does not need to converge to the most efficient outcome. Instead, players may be trapped in local optima of the fitness landscape, in which players only partly invest into the public good[79]. These simulations suggest that players sometimes confine themselves to be partial partners: in equilibrium, they contribute a considerable amount of their endowment to the public good, but they may not contribute everything. To obtain a general understanding when full partnerships evolve, a complete characterization of all partner and rival strategies for games with multiple actions would be desirable.

Conclusion

Direct reciprocity is a mechanism for the evolution cooperation. It is based on repeated interactions between the same individuals. The new mathematical formalism of ZD strategies has led to a characterization of evolutionarily successful strategies into partners and rivals. Partners aim for mutual cooperation, but are ready to defend themselves when being exploited. Rivals focus on their own relative advantage and on winning. Only partner strategies stabilize cooperation.The rivals' aim to put themselves first, which is a widespread motivation of current populist politics, ensures a path toward destruction.

Methods

For the simulation results shown in **Fig. 5** and **Fig. S2** we have used the method proposed by Imhof & Nowak[33]. We consider a population of size N, which initially consists of ALLD players only. At each time step, one individual is chosen to experiment with a new strategy. This mutant strategy is generated by randomly drawing the cooperation probabilities from the interval [0,1]. If the mutant strategy yields a payoff of $\pi_M(j)$, where j is the number of mutants in the population, and if residents get a payoff of $\pi_R(j)$, then the fixation probability ρ of the mutant strategy is[97]

$$\rho = \left(1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} \exp\left[-s\big(\pi_M(j) - \pi_R(j)\big)\right]\right)^{-1}. \tag{1}$$

The parameter $s \geq 0$ measures the strength of selection. If $s = 0$ payoffs are irrelevant and the fixation probability simplifies to $\rho = 1/N$. For larger values of s, the evolutionary process increasingly favors the fixation of strategies that yield high payoffs. Once the mutant strategy has either reached fixation, or gone to extinction, another mutant strategy is introduced.

We have iterated this process for 10^7 mutant strategies per simulation run. The process approximates the evolutionary dynamics of finite populations when mutations are sufficiently rare[98,99]. It generates a sequence $(\mathbf{p}_0, \mathbf{p}_1, \ldots)$, where $\mathbf{p}_t$ is the strategy the residents apply at time t. Using this sequence, we can compute how often players cooperate on average, and how often they apply an approximate partner or rival strategy. We compare the abundance of these strategies with their abundance under neutral evolution, $s = 0$, in which case the abundance coincides with the volume of these strategy sets.

References

[1] von Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, 1944).

[2] Nash, J. F. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences USA* **36**, 48–49 (1950).

[3] Flood, M. M. Some experimental games. *Management Science* **5**, 5–26 (1958).

[4] Rapoport, A. & Chammah, A. M. *Prisoner's Dilemma* (University of Michigan Press, Ann Arbor, 1965).

[5] Trivers, R. L. The evolution of reciprocal altruism. *The Quarterly Review of Biology* **46**, 35–57 (1971).

[6] Nowak, M. A. *Evolutionary dynamics* (Harvard University Press, Cambridge MA, 2006).

[7] Sigmund, K. *The Calculus of Selfishness* (Princeton Univ. Press, 2010).

[8] Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).

[9] Smale, S. The prisoner's dilemma and dynamical systems associated to non-cooperative games. *Econometrica* **48**, 1617–1634 (1980).

[10] Mailath, G. J. & Samuelson, L. *Repeated games and reputations* (Oxford Univ. Press, Oxford, UK, 2006).

Extensive compendium on repeated games from an economics point of view, which gives an excellent overview on the Folk theorem literature.

[11] Abreu, D. Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory* **39**, 191–225 (1986).

[12] Bernheim, D. & Whinston, M. D. Multimarket contact and collusive behavior. *The RAND journal of economics* **21**, 1–26 (1990).

[13] Cable, D. M. & Shane, S. A prisoner's dilemma approach to entrepreneur-venture capitalist relationships. *Academy of Management Review* **22**, 142–176 (1997).

[14] Majeski, S. J. Arms races as iterated prisoner's dilemma games. *Mathematical Social Sciences* **7**, 253–266 (1984).

[15] Aumann, R. J. War and peace. *Proceedings of the National Academy of Sciences USA* **103**, 17075–17078 (2006).

[16] Wilkinson, G. S. Reciprocal food-sharing in the vampire bat. *Nature* **308**, 181–184 (1984).

[17] Stephens, D. W., McLinn, C. M. & Stevens, J. R. Discounting and reciprocity in an iterated prisoner's dilemma. *Science* **298**, 2216–2218 (2002).

[18] Milinski, M. Tit For Tat in sticklebacks and the evolution of cooperation. *Nature* **325**, 433–435 (1987).

[19] Ben-Porath, E. The complexity of computing a best response automaton in repeated games with mixed strategies. *Games and Economic Behavior* **2**, 1–12 (1990).

[20] Papadimitriou, C. H. On players with a bounded number of states. *Games and Economic Behavior* **4**, 122–131 (1992).

[21] Friedman, J. A non-cooperative equilibrium for supergames. *Review of Economic Studies* **38**, 1–12 (1971).

[22] Aumann, R. J. Survey of repeated games. In Henn, R. & Moeschlin, O. (eds.) *Essays in game theory and mathematical economics in honor of Oskar Morgenstern*, chap. Survey of repeated games (Wissenschaftsverlag, 1981).

[23] Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, 1982).

[24] Selten, R. & Hammerstein, P. Gaps in harley's argument on evolutionarily stable learning rules and in the logic of "Tit for Tat". *Behavioral and Brain Sciences* **7**, 115–116 (1984).

[25] Boyd, R. & Lorberbaum, J. No pure strategy is evolutionary stable in the iterated prisoner's dilemma game. *Nature* **327**, 58–59 (1987).

[26] Bendor, J. & Swistak, P. Types of evolutionary stability and the problem of cooperation. *Proceedings of the National Academy of Sciences USA* **92**, 3596–3600 (1995).

[27] van Veelen, M., García, J., Rand, D. G. & Nowak, M. A. Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences USA* **109**, 9929–9934 (2012).

[28] Garcia, J. & van Veelen, M. In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory* **161**, 161–189 (2016).

[29] Boyd, R. Mistakes allow evolutionary stability in the repeated Prisoner's Dilemma game. *Journal of Theoretical Biology* **136**, 47–56 (1989).

[30] Fudenberg, D. & Maskin, E. Evolution and cooperation in noisy repeated games. *American Economic Review* **80**, 274–279 (1990).

[31] Nowak, M. A. & Sigmund, K. Chaos and the evolution of cooperation. *Proceedings of the National Academy of Sciences USA* **90**, 5091–5094 (1993).

[32] Imhof, L. A., Fudenberg, D. & Nowak, M. A. Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences USA* **102**, 10797–10800 (2005).

[33] Imhof, L. A. & Nowak, M. A. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B* **277**, 463–468 (2010).

[34] Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428, 646–650 (2004).
Introduces finite population size to evolutionary game dynamics.

[35] Imhof, L. A., Fudenberg, D. & Nowak, M. A. Tit-for-tat or win-stay, lose-shift? *Journal of Theoretical Biology* **247**, 574–580 (2007).

[36] García, J. & Traulsen, A. The structure of mutations and the evolution of cooperation. *PLoS One* **7**, e35287 (2012).

[37] Kurokawa, S. & Ihara, Y. Emergence of cooperation in public goods games. *Proceedings of the Royal Society B* **276**, 1379–1384 (2009).

[38] Martinez-Vaquero, L. A., Cuesta, J. A. & Sanchez, A. Generosity pays in the presence of direct reciprocity: A comprehensive study of 2x2 repeated games. *PLoS ONE* **7**, E35135 (2012).

[39] Baek, S. K., Jeong, H. C., Hilbe, C. & Nowak, M. A. Comparing reactive and memory-one strategies of direct reciprocity. *Scientific Reports* **6**, 25676 (2016).

[40] Axelrod, R. *The evolution of cooperation* (Basic Books, New York, NY, 1984).
Axelrod's tournament and the success of Tit-for-Tat have been transformative for the field; this book contains a detailed analysis of the tournament's results.

[41] Molander, P. The optimal level of generosity in a selfish, uncertain environment. *Journal of Conflict Resolution* **29**, 611–618 (1985).

[42] Nowak, M. A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992).

[43] Nowak, M. A. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* 364, 56–58 (1993).

Describes a simple yet surprisingly powerful strategy to maintain cooperation in noisy repeated games, Win-stay Lose-shift.

[44] Kraines, D. P. & Kraines, V. Y. Pavlov and the prisoner's dilemma. *Theory and Decision* **26**, 47–79 (1989).

[45] Lindgren, K. Evolutionary dynamics in game-theoretic models. In Arthur, W. B., Durlauf, S. N. & Lane, D. A. (eds.) *The Economy as an Evolving Complex System II* (Addison-Wesley, 1997).

[46] Hauert, C. & Schuster, H. G. Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. *Proceedings of the Royal Society B* **264**, 513–519 (1997).

[47] Pinheiro, F. L., Vasconcelos, V. V., Santos, F. C. & Pacheco, J. M. Evolution of all-or-none strategies in repeated public goods dilemmas. *PLoS Comput Biol* **10**, e1003945 (2014).

[48] Hilbe, C., Martinez-Vaquero, L. A., Chatterjee, K. & Nowak, M. A. Memory-n strategies of direct reciprocity. *Proceedings of the National Academy of Sciences USA* **114**, 4715–4720 (2017).

[49] Fischer, I. *et al.* Fusing enacted and expected mimicry generates a winning strategy that promotes the evolution of cooperation. *PNAS* **110**, 10229–10233 (2013).

[50] Yi, S. D., Baek, S. K. & Choi, J.-K. Combination with anti-tit-for-tat remedies problems of tit-for-tat. *Journal of Theoretical Biology* **412**, 1–7 (2017).

[51] Duersch, P., Oechssler, J. & Schipper, B. When is tit-for-tat unbeatable? *International Journal of Game Theory* **43**, 25–36 (2013).

[52] Rapoport, A., Seale, D. A. & Colman, A. M. Is Tit-for-Tat the answer? On the conclusions drawn from axelrod's tournaments. *PLoS One* **10**, e0134128 (2015).

[53] Bendor, J. In good times and bad: Reciprocity in an uncertain world. *American Journal of Political Science* **31**, 531–558 (1987).

[54] Nowak, M. A., Sigmund, K. & El-Sedy, E. Automata, repeated games and noise. *Journal of Mathematical Biology* **33**, 703–722 (1995).

[55] Brandt, H. & Sigmund, K. The good, the bad and the discriminator - errors in direct and indirect reciprocity. *Journal of Theoretical Biology* **239**, 183–194 (2006).

[56] Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, UK, 1998).

[57] Cressman, R. *Evolutionary Dynamics and Extensive Form Games* (MIT Press, Cambridge, 2003).

[58] Nowak, M. A. & Sigmund, K. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematicae* **20**, 247–265 (1990).

[59] Press, W. H. & Dyson, F. D. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *PNAS* 109, 10409–10413 (2012).

Using innovative mathematical concepts, Press & Dyson show there are strategies for the repeated prisoner's dilemma that allow players to extort their opponents.

[60] Chen, J. & Zinger, A. The robustness of zero-determinant strategies in iterated prisoner's dilemma games. *Journal of Theoretical Biology* **357**, 46–54 (2014).

[61] Pan, L., Hao, D., Rong, Z. & Zhou, T. Zero-determinant strategies in iterated public goods game. *Scientific Reports* **5**, 13096 (2015).

[62] McAvoy, A. & Hauert, C. Autocratic strategies for iterated games with arbitrary action spaces. *Proceedings of the National Academy of Sciences* **113**, 3573–3578 (2016).

[63] McAvoy, A. & Hauert, C. Autocratic strategies for alternating games. *Theoretical Population Biology* **113**, 13–22 (2016).

[64] Ichinose, G. & Masuda, N. Zero-determinant strategies in finitely repeated games. *Journal of Theoretical Biology* **438**, 61–77 (2018).

[65] Hilbe, C., Nowak, M. A. & Sigmund, K. The evolution of extortion in iterated prisoner's dilemma games. *Proceedings of the National Academy of Sciences USA* 110, 6913–6918 (2013).
Based on computer simulations, this article suggests that extortionate strategies can only succeed in small populations, or when two populations evolve at different rates.

[66] Hilbe, C., Nowak, M. A. & Traulsen, A. Adaptive dynamics of exortion and compliance. *PLoS One* **8**, e77886 (2013).

[67] Szolnoki, A. & Perc, M. Evolution of extortion in structured populations. *Physical Review E* **89**, 022804 (2014).

[68] Szolnoki, A. & Perc, M. Defection and extortion as unexpected catalysts of unconditional cooperation in structured populations. *Scientific Reports* **4**, 5496 (2014).

[69] Adami, C. & Hintze, A. Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything. *Nature Communications* **4**, 2193 (2013).

[70] Wu, Z.-X. & Rong, Z. Boosting cooperation by involving extortion in spatial prisoner's dilemma games. *Physical Review E* **90**, 062102 (2014).

[71] Hilbe, C., Röhl, T. & Milinski, M. Extortion subdues human players but is finally punished in the prisoner's dilemma. *Nature Communications* **5**, 3976 (2014).

[72] Xu, B., Zhou, Y., Lien, J. W., Zheng, J. & Wang, Z. Extortion can outperform generosity in iterated prisoner's dilemma. *Nature Communications* **7**, 11125 (2016).

[73] Milinski, M., Hilbe, C., Semmann, D., Sommerfeld, R. D. & Marotzke, J. Humans choose representatives who enforce cooperation in social dilemmas through extortion. *Nature Communications* **7**, 10915 (2016).

[74] Hilbe, C., Hagel, K. & Milinski, M. Asymmetric power boosts extortion in an economic experiment. *PLoS ONE* **11**, e0163867 (2016).

[75] Hilbe, C., Wu, B., Traulsen, A. & Nowak, M. A. Cooperation and control in multiplayer social dilemmas. *Proceedings of the National Academy of Sciences USA* **111**, 16425–16430 (2014).

[76] Stewart, A. J. & Plotkin, J. B. From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proceedings of the National Academy of Sciences USA* 110, 15348–15353 (2013).
Shows that large evolving populations favor the emergence of generous strategies, and introduces the important concept of evolutionary robustness.

[77] Stewart, A. J. & Plotkin, J. B. Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences USA* 111, 17558 – 17563 (2014).
Describes all evolutionary robust strategies for iterated 2×2 games.

[78] Stewart, A. J. & Plotkin, J. B. Small groups and long memories promote cooperation. *Scientific Reports* **6**, 26889 (2016).

[79] Stewart, A. J., Parsons, T. L. & Plotkin, J. B. Evolutionary consequences of behavioral diversity. *Proceedings of the National Academy of Sciences USA* **113**, E7003–E7009 (2016).

[80] Akin, E. The iterated prisoner's dilemma: Good strategies and their dynamics. In Assani, I. (ed.) *Ergodic Theory, Advances in Dynamics*, 77–107 (de Gruyter, 2016).
Based on the mathematical formalism of ZD strategies, Akin was first to characterize all memory-1 partner strategies (called "good strategies" in this article).

[81] Akin, E. What you gotta know to play good in the iterated prisoner's dilemma. *Games* **6**, 175–190 (2015).

[82] Akin, E. Good strategies for the iterated prisoner's dilemma: Smale vs. Markov. *Journal of Dynamics and Games* **4**, 217–253 (2017).

[83] Hilbe, C., Traulsen, A. & Sigmund, K. Partners or rivals? Strategies for the iterated prisoner's dilemma. *Games and Economic Behavior* 92, 41–52 (2015).
Introduces the notion of competitive rival strategies, and describes partner and rival strategies for the repeated prisoner's dilemma with discounted payoffs.

[84] van den Berg, P., Molleman, L. & Weissing, F. J. Focus on the success of others leads to selfish behavior. *Proceedings of the National Academy of Sciences USA* **112**, 2912–2917 (2015).

[85] Dal Bó, P. Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *American Economic Review* **95**, 1594–1604 (2005).

[86] Blonski, M., Ockenfels, P. & Spagnolo, G. Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence. *American Economic Journal: Microeconomics* **3**, 164–192 (2011).

[87] Fudenberg, D., Dreber, A. & Rand, D. G. Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review* **102**, 720–749 (2012).

[88] Doebeli, M. & Hauert, C. Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecology Letters* **8**, 748–766 (2005).

[89] Nowak, M. A. Evolving cooperation. *Journal of Theoretical Biology* **299**, 1–8 (2012).

[90] Frean, M. R. The prisoner's dilemma without synchrony. *Proceedings of the Royal Society B* **257**, 75–79 (1994).

[91] Nowak, M. A. & Sigmund, K. The alternating prisoner's dilemma. *Journal of Theoretical Biology* **168**, 219–226 (1994).

[92] Zagorsky, B. M., Reiter, J. G., Chatterjee, K. & Nowak, M. A. Forgiver triumphs in alternating prisoner's dilemma. *PLoS One* **8**, e80814 (2013).

[93] Roberts, G. & Sherratt, T. N. Development of cooperative relationships through increasing investment. *Nature* **394**, 175–179 (1998).

[94] Wahl, L. M. & Nowak, M. A. The continuous prisoner's dilemma: I. Linear reactive strategies. *Journal of Theoretical Biology* **200**, 307–321 (1999).

[95] Killingback, T. & Doebeli, M. The continuous Prisoner's Dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *The American Naturalist* **160**, 421–438 (2002).

[96] Gokhale, C. S. & Traulsen, A. Evolutionary games in the multiverse. *Proceedings of the National Academy of Sciences USA* **107**, 5500–5504 (2010).

[97] Traulsen, A., Nowak, M. A. & Pacheco, J. M. Stochastic dynamics of invasion and fixation. *Physical Review E* **74**, 011909 (2006).

[98] Fudenberg, D. & Imhof, L. A. Imitation processes with small mutations. *Journal of Economic Theory* **131**, 251–262 (2006).

[99] Wu, B., Gokhale, C. S., Wang, L. & Traulsen, A. How small are small mutation rates? *Journal of Mathematical Biology* **64**, 803–827 (2012).

[100] Boerlijst, M. C., Nowak, M. A. & Sigmund, K. Equal pay for all prisoners. *American Mathematical Monthly* **104**, 303–307 (1997).

Acknowledgments. This work was supported by the European Research Council Start Grant 279307: Graph Games (to K.C.), Austrian Science Fund (FWF) Grant P23499-N23 (to K.C.), FWF NFN Grant S11407-N23 Rigorous Systems Engineering/Systematic Methods in Systems Engineering (to K.C.), Office of Naval Research Grant N00014-16-1- 2914 (to M.A.N.), and the John Templeton Foundation (M.A.N.). C.H. acknowledges generous support from the IST-FELLOW program.

Author contributions. All authors conceived the study, performed the analysis, and wrote the manuscript.

Author information. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.H. (christian.hilbe@ist.ac.at).

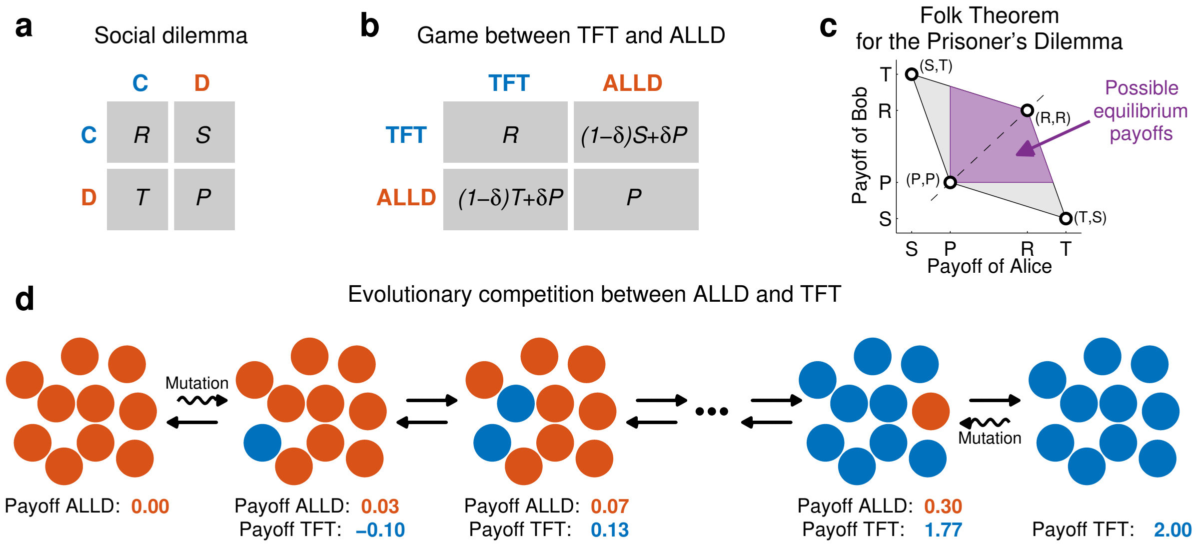

Figure 1: Repeated interactions allow evolution of cooperation. a, In a social dilemma, two cooperators get a higher payoff than two defectors, $R > P$, but there is a temptation to defect. The temptation can come in three forms: $T > R$, $P > S$, or $T > S$. The game is a social dilemma if at least one of those inequalities holds. The prisoner's dilemma (PD) is the most stringent social dilemma; here all three temptations hold. The PD is defined by the payoff ranking $T > R > P > S$. **b,** If the PD is repeated with probability δ, players can use conditionally cooperative strategies like Tit-for-Tat (TFT). TFT yields the mutual cooperation payoff R against itself, and it is stable against ALLD if δ is sufficiently large. **c,** The Folk theorem states that for sufficiently large δ, all payoff pairs in which both players get at least P can arise in equilibrium. **d,** In stochastic evolutionary dynamics TFT can invade ALLD. A single TFT mutant can have a fixation probability that exceeds the neutral probability $1/N$, where N is the population size[34]. Parameters: $N = 10$, $R = 2$, $S = -1$, $T = 3$, $P = 0$, and $\delta = 0.9$.

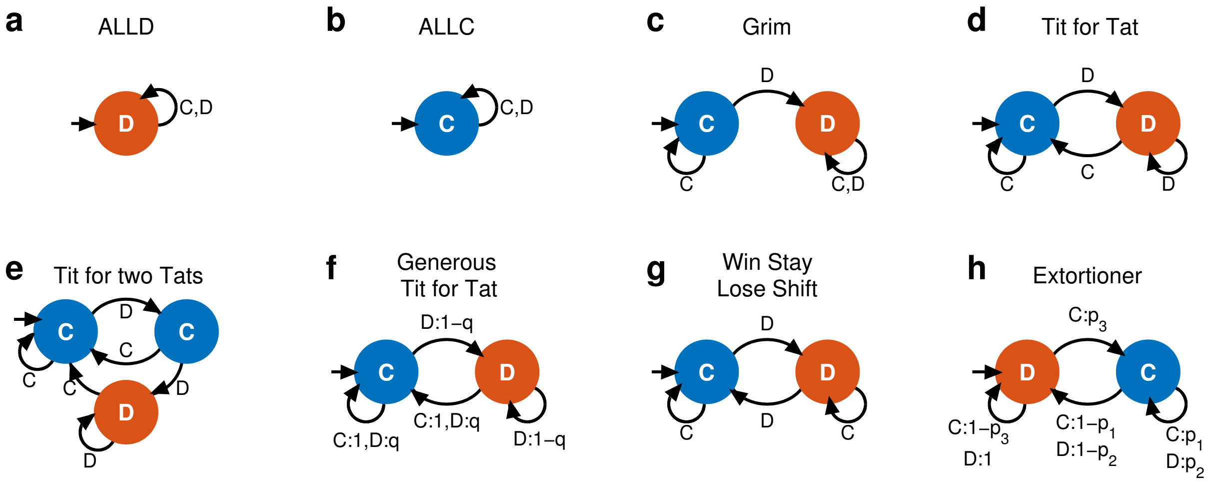

Figure 2: Eight strategies for the repeated prisoner's dilemma. Each strategy is shown as a finite state automaton[54]. The colored vertices indicate the player's next action. The arrows represent transitions between states after each round. The black letters C and D represent the co-player's action. The arrow from the left points at the initial state. **a,** ALLD always defects. **b,** ALLC always cooperates. **c,** Grim cooperates until the co-player defects once, then it defects forever. **d,** Tit for tat (TFT) cooperates in the first round, then repeats what the co-player did in the previous round. **e,** Tit for two tats (TF2T) is similar to TFT, but it takes two consecutive defections of the co-player for TF2T to retaliate. **f,** Generous Tit for Tat cooperates in the first round and if the co-player has cooperated in the previous round; it cooperates with probability $q < q^*$ if the co-player has defected. The threshold q^* ensures that no other strategy can invade (**Box 2**). **g,** Win Stay Lose Shift cooperates in the first round, and it repeats its own move if the payoff was T or R; it switches to the other move if the payoff was P or S. **h,** An extortioner defects in the first round; then defects if both players have defected; cooperates with probabilities p_1, p_2, p_3 if the previous round was CC, CD, or DC. These probabilities are chosen such that the two players' payoffs are on a line (**Fig. 3a**). TF2T requires the player to remember the outcome of the past two rounds; all other depicted strategies are memory-1 strategies.

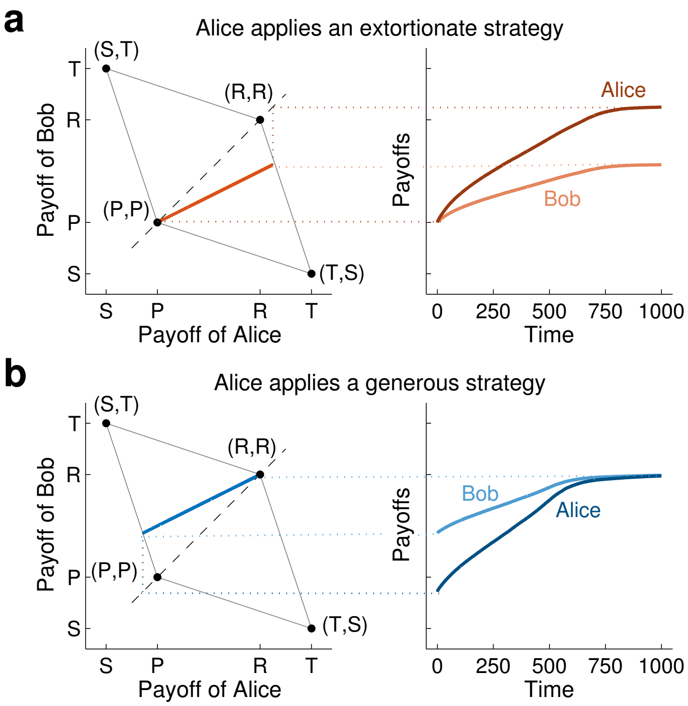

Figure 3: Adaptive players versus zero-determinant (ZD) strategies. a, If Alice applies an extortionate ZD strategy, she enforces a linear relationship between her and Bob's payoff (left). Which point on that line will be realized, depends on Bob's strategy. But because the red line in the left panel is on or below the main diagonal, Alice never gets a lower payoff than Bob. Moreover, the slope of the line is positive: if Bob adapts his strategy to improve his own payoff, he also improves Alice's payoff. In the long run, Alice's payoff is not only higher than Bob's, but also higher than the mutual cooperation payoff (right). **b,** If Alice applies a generous ZD strategy[76], she also enforces a positive relation between the players' payoffs, but now Bob's payoff is higher than Alice's. However, as Bob adapts to improve his payoff, both players move towards mutual cooperation.

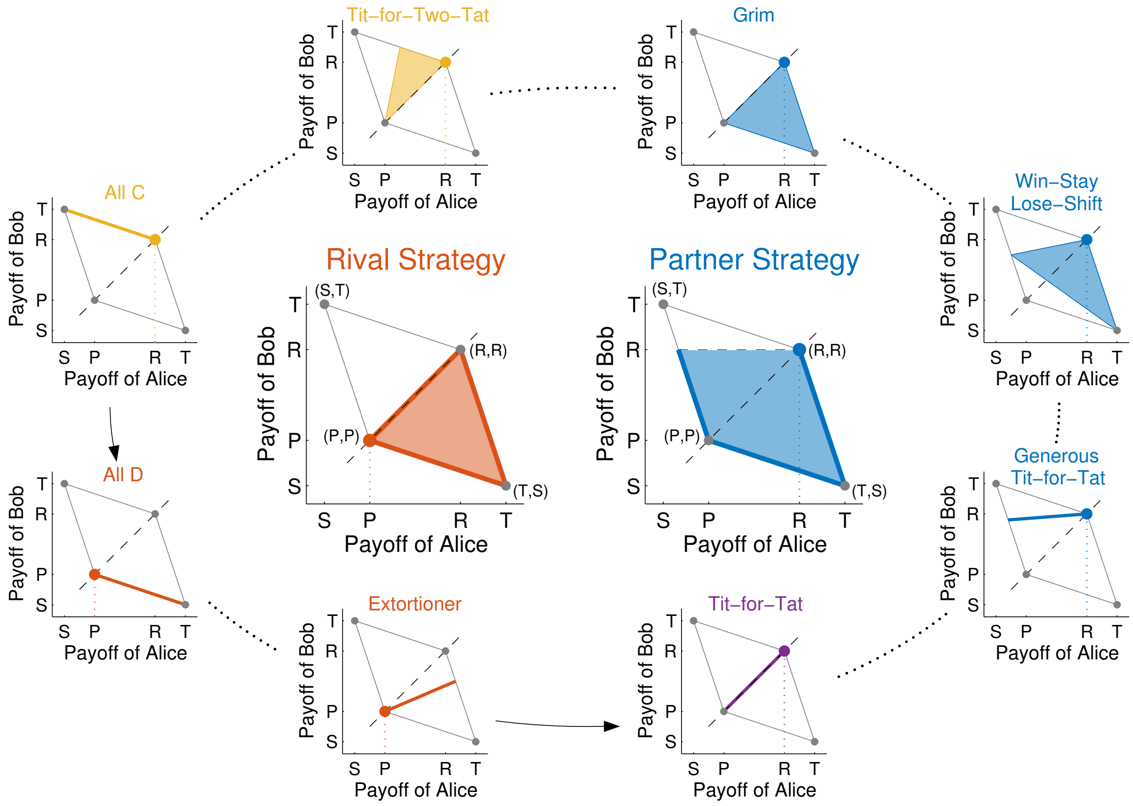

Figure 4: Partners and rivals. In each panel, the grey diamond depicts the space of possible payoffs for the two players. The colored areas or lines in the periphery show the feasible payoffs when Alice uses ALLD, Extortion, TFT, GTFT, WSLS, Grim, TF2T, or ALLC. The colored dot denotes the payoff when Bob uses the same strategy as Alice. Most of these strategies either qualify as rival (red) or partner (blue). With a rival strategy, Alice can outperform her opponent; irrespective of Bob's strategy, she always obtains at least the payoff of Bob. With a partner strategy, Alice aims to reach the mutual cooperation payoff without tolerating exploitation. In that case, Bob may be able to get a larger payoff than Alice, but he cannot get a larger payoff than R. The payoff relations correspond to the infinitely repeated game without discounting. In that case, TFT is both a rival and a partner. Payoffs are $R = b - c$, $S = -c$, $T = b$ and $P = 0$ with $c = 1$ and $b = 3$.

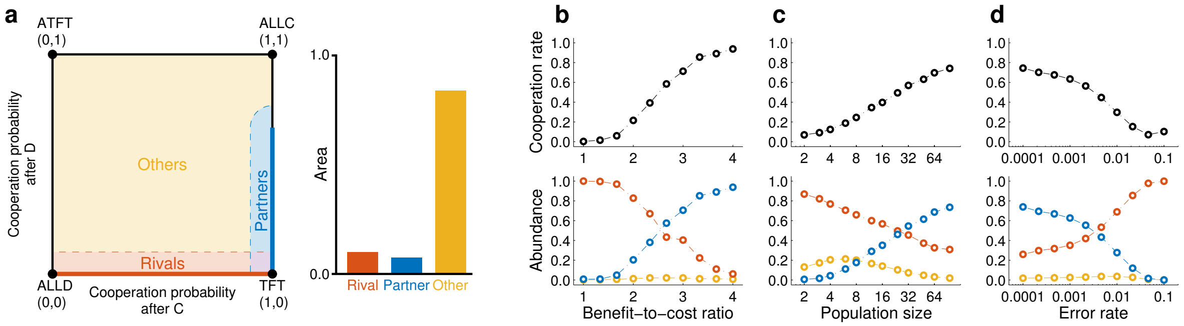

Figure 5: Evolution favors partners or rivals. **a,** For the infinitely repeated Prisoner's Dilemma, reactive strategies are points (p, q) in the unit square: p and q are the probabilities to cooperate given the co-player has cooperated or defected in the previous round. Rivals are given by $q = 0$ (red boundary). Partners are given by $p = 1$ and $q < q^*$, with q^* as defined in **Box 2** (blue boundary). A reactive strategy is an ε-rival, or an ε-partner, if its distance to the respective set of strategies is at most ε. We show the area of ε-rivals (red) and ε-partners (blue) for $\varepsilon = 0.1$. For reactive strategies, the area of ε-partners depends on the payoffs, but it is always smaller than the area of ε-rivals. Most strategies are neither partners nor rivals (yellow). **b–d,** To explore the evolutionary relevance of partners and rivals, we have simulated a pairwise comparison process in a finite population (see **Methods**). We vary three parameters: efficiency of cooperation, population size, and error rate. The upper panel shows the resulting cooperation rates. The lower panel gives the fraction of time the population uses ε-rivals, ε-partners, or other strategies. Parameters: $T = b$, $R = b - c$, $P = 0$ and $S = -c$, where $c = 1$ is the cost of cooperation and b the benefit to the co-player. Unless stated otherwise, $b = 3$, $N = 50$, and error rate is zero.

Box 1: How to gain twice as much as your opponent.

Assume that Alice and Bob interact in a repeated Prisoner's Dilemma (PD) with payoffs $T > R > P > S$. For simplicity, we assume there is always another round, $\delta = 1$. Alice uses a memory-1 strategy (p_1, p_2, p_3, p_4). Bob uses an arbitrary strategy. Suppose that over the course of the entire game between Alice and Bob, the four outcomes CC, CD, DC, DD occur with relative frequencies v_1, v_2, v_3, v_4. Alice's probability to switch from cooperation to defection is $(1-p_1)v_1 + (1-p_2)v_2$. Her probability to switch from defection to cooperation is $p_3v_3 + p_4v_4$. Because Alice can only switch from cooperation to defection if she has switched from defection to cooperation before, we obtain Akin's identity[80,81]

$$(1-p_1)v_1 + (1-p_2)v_2 = p_3v_3 + p_4v_4. \tag{2}$$

Let us assume Alice uses a particular rule to determine the four probabilities of her memory-1 strategy. She chooses three constants α, β, and γ and then takes the four probabilities

$$\begin{aligned} p_1 &= \alpha R + \beta R + \gamma + 1, \\ p_2 &= \alpha S + \beta T + \gamma + 1, \\ p_3 &= \alpha T + \beta S + \gamma, \\ p_4 &= \alpha P + \beta P + \gamma. \end{aligned} \tag{3}$$

Such a strategy is called zero-determinant (ZD) strategy[59]. From (2) and (3) we obtain

$$\alpha(Rv_1+Sv_2+Tv_3+Pv_4) + \beta(Rv_1+Tv_2+Sv_3+Pv_4) + \gamma = 0. \tag{4}$$

The expression $Rv_1+Sv_2+Tv_3+Pv_4$ is exactly Alice's payoff for the repeated game, π_A. Similarly, $Rv_1+Tv_2+Sv_3+Pv_4$ is Bob's payoff, π_B. Thus, if Alice applies a ZD strategy, the payoffs of the two players satisfy

$$\alpha\pi_A + \beta\pi_B + \gamma = 0. \tag{5}$$

Curiously, the values of α, β, γ are solely determined by Alice. In the following, let us assume that the payoffs are normalized such that $P=0$. Then Alice may set $\alpha=-\beta/2$ and $\gamma=0$. The remaining parameter β she can choose arbitrarily, subject to the restriction that $\beta \neq 0$ and that the four probabilities in (3) satisfy $0 \leq p_i \leq 1$. In that case Eq. (5) simplifies to

$$\pi_A = 2\pi_B. \tag{6}$$

Alice earns twice as much as Bob, irrespective of Bob's strategy. Moreover, if Bob tries to increase his own payoff by using another strategy, he simultaneously always increases Alice's payoff, too.

ZD strategies with $\alpha=-\chi\beta$ and $\gamma=\beta(\chi-1)P$ are called *extortionate*. The parameter χ with $0<\chi<1$ determines by how much Alice's payoff exceeds Bob's. For $P \neq 0$, extortionate strategies enforce $(\pi_A - P) = \frac{1}{\chi}(\pi_B - P)$. That is, if the two players get a payoff higher than P, Alice gets a disproportionate share of the surplus.

Some ZD strategies have been known before. For example, if Alice uses an 'equalizer' strategy, she imposes a fixed payoff for Bob irrespective of Bob's strategy.[100]

Box 2: Of partners and rivals.

When Alice and Bob play a repeated Prisoner's Dilemma (PD) with $T+S<2R$, Alice applies a *partner strategy* (called *good strategy* by Akin[80,81]) if the following two conditions hold:

(*i*) If Bob applies the same strategy as Alice, both get the mutual cooperation payoff, $\pi_A=\pi_B=R$.

(*ii*) By applying a different strategy, Bob can get at most R, in which case Alice gets the same payoff. That is, if $\pi_B \geq R$ then $\pi_B=\pi_A=R$.

In contrast, Alice applies a *rival strategy* (or *competitive strategy*[83]) if she always gets at least the payoff of Bob, $\pi_A \geq \pi_B$. The two definitions make no restriction on Bob's strategy. Bob may remember arbitrarily many rounds.

We can characterize all partners and rivals among reactive strategies (y, p, q). Without discounting, $\delta=1$, a reactive strategy is a partner if and only if $p=1$ and $q<q^*$ with $q^*=\min\{1-(T-R)/(R-S),\ (R-P)/(T-P)\}$. It is a rival if and only if $q=0$. In both cases, the initial cooperation probability y can be chosen arbitrarily; the only exception are strategies with $p=1$ and $q=0$. Such strategies are always rivals, but TFT with $y=1$ is the only such strategy that is also a partner.

A similar characterization is possible for discounted games and for memory-1 strategies[83]. In that case, Alice's strategy $(p_0;\ p_1, p_2, p_3, p_4)$ is a partner if and only if

$$\begin{aligned} p_0 = p_1 &= 1, \\ \delta(T-R)p_4 - \delta(R-P)(1-p_2) + (1-\delta)(T-R) &< 0, \\ \delta(T-R)p_3 - \delta(R-S)(1-p_2) + (1-\delta)(T-R) &< 0. \end{aligned} \tag{7}$$

A partner strategy is never the first to defect. TFT is a partner strategy if $\delta>(T-R)/(T-P)$ and $\delta>(T-R)/(R-S)$. WSLS is a partner strategy if $\delta>(T-R)/(R-P)$ and $\delta>(T-R)/(T-S)$, which is a sharper condition. Alice uses a rival strategy if

$$\begin{gathered} p_1 \text{ arbitrary}, \\ p_0 = p_4 = 0, \\ \delta(p_2+p_3) \leq 1. \end{gathered} \tag{8}$$

ALLD is always a rival strategy. Extortion is a rival strategy if $2P < T+S$.

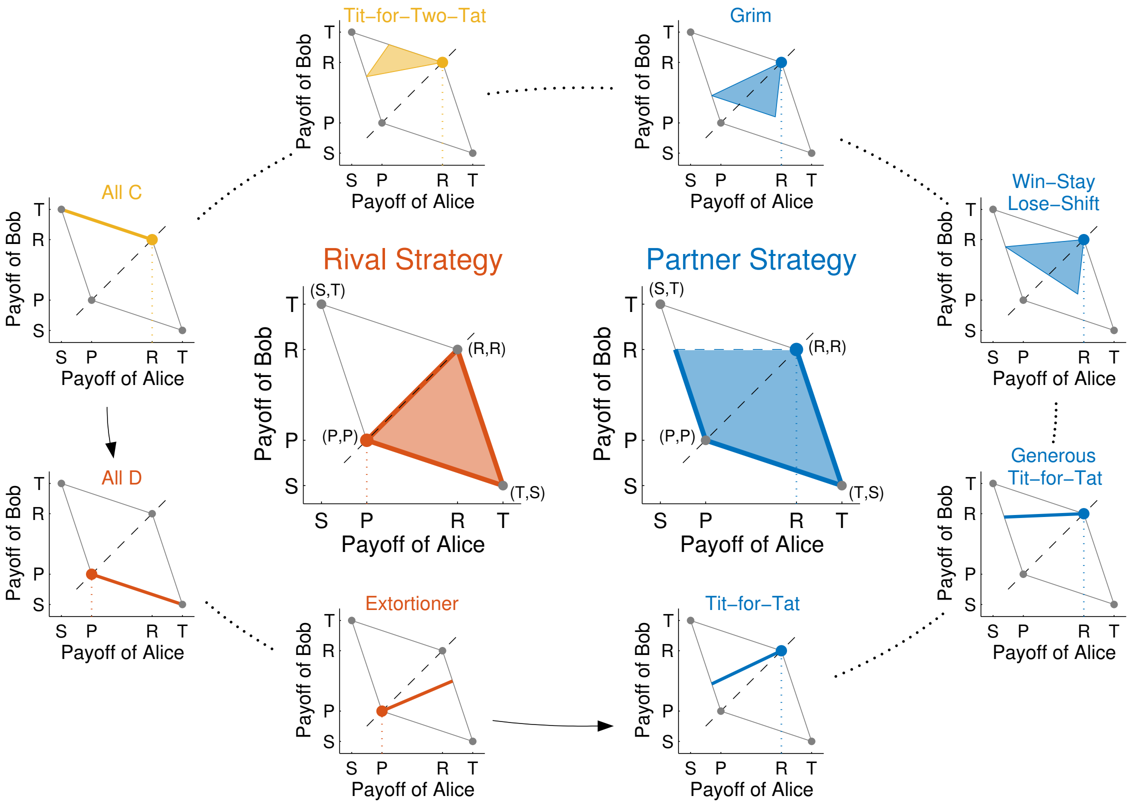

Figure S1: Partners and rivals in games with discounted payoffs. We have calculated which payoffs are feasible for the eight strategies in **Fig. 4** when the game only has a finite expected number of rounds. The payoff relationships for ALLC and ALLD remain unchanged. For the depicted parameter value $\delta = 0.7$, we can redefine the memory-1 strategies for GTFT and for the extortionate strategy, such that they enforce the same payoff relationship as in **Fig. 4**. For the other four strategies, TFT, Grim, WSLS, TF2T, the feasible payoffs change. In particular, when the expected number of rounds is finite, TFT is a partner strategy but not a rival strategy (since TFT cooperates in the very first round). Parameters are the same as in **Fig. 4**, except for δ.

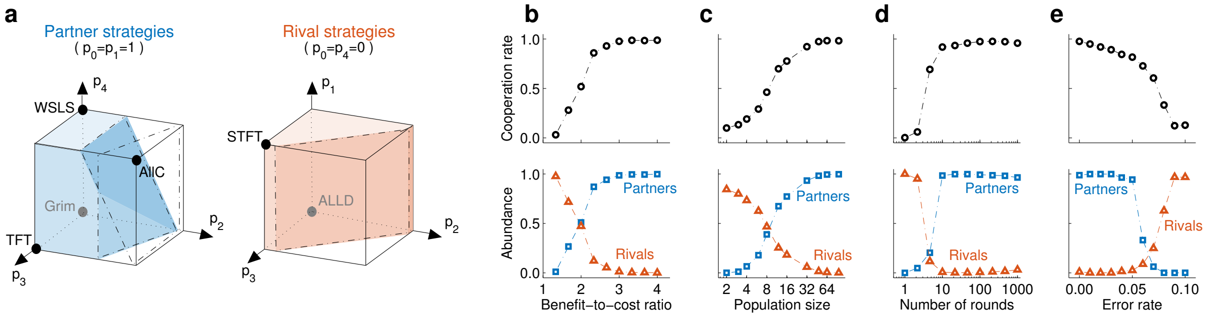

Figure S2: Evolution of partners and rivals among memory-1 strategies. We have explored the evolution of partner and rival strategies when players use memory-1 strategies $(p_0;\ p_1, p_2, p_3, p_4)$. **a,** In repeated games with discounting, $\delta < 1$, partner strategies are required to set $p_0 = p_1 = 1$ to allow for mutual cooperation. The other three elements p_2, p_3, p_4 need to obey the two inequalities in (7). Rivals are required to set $p_0 = p_4 = 0$ to ensure that no opponent can outperform them in a pairwise encounter. The value of p_1 can be chosen arbitrarily, while the values of p_2 and p_3 need to satisfy $\delta(p_2 + p_3) \leq 1$, see Eq. (8). Both sets have measure zero within the set of all memory-1 strategies. **b–e** We have simulated a pairwise imitation process on the space of memory-1 strategies, and we have recorded the average cooperation rate (upper panel) as well as the frequency with which players use an approximate partner or rival strategy. Here, we define a memory-1 strategy to be an approximate partner strategy if it yields a payoff of at least $R(1-\varepsilon)$ against itself, and if the inequalities in (7) are satisfied. We speak of an approximate rival strategy if it yields a payoff of at most $P(1+\varepsilon)$ against itself, and if the inequality in (8) is satisfied. We use $\varepsilon = 0.2$, for which approximate partners and rivals make up roughly 5% of the volume of all memory-1 strategies. Approximate partners are favored when cooperation yields a high benefit, when the population is large, when there is a substantial number of round, and when actions are implemented reliably. Parameters: For the simulations we use payoffs $R = b - c$, $S = -c$, $T = b$ and $P = 0$ with $b = 3$ and $c = 1$, population size $N = 50$, $\delta = 0.995$ (corresponding to an expected number of 200 rounds), error rate $\varepsilon = 0$, and selection strength $s = 10$. In **a**, we use $\delta = 0.9$ for better clarity.