LETTER

https://doi.org/10.1038/s41586-018-0277-x

Evolution of cooperation in stochastic games

Christian Hilbe[1,2]*, Štěpán Šimsa[3], Krishnendu Chatterjee[2]* & Martin A. Nowak[1,4]*

Social dilemmas occur when incentives for individuals are misaligned with group interests[1–7]. According to the 'tragedy of the commons', these misalignments can lead to overexploitation and collapse of public resources. The resulting behaviours can be analysed with the tools of game theory[8]. The theory of direct reciprocity[9–15] suggests that repeated interactions can alleviate such dilemmas, but previous work has assumed that the public resource remains constant over time. Here we introduce the idea that the public resource is instead changeable and depends on the strategic choices of individuals. An intuitive scenario is that cooperation increases the public resource, whereas defection decreases it. Thus, cooperation allows the possibility of playing a more valuable game with higher payoffs, whereas defection leads to a less valuable game. We analyse this idea using the theory of stochastic games[16–19] and evolutionary game theory. We find that the dependence of the public resource on previous interactions can greatly enhance the propensity for cooperation. For these results, the interaction between reciprocity and payoff feedback is crucial: neither repeated interactions in a constant environment nor single interactions in a changing environment yield similar cooperation rates. Our framework shows which feedbacks between exploitation and environment—either naturally occurring or designed—help to overcome social dilemmas.

The tragedy of the commons leads to the question of how to manage and conserve public resources[1–8]. Any solution to this problem requires an understanding of which processes drive human cooperation and how institutions, norms and other feedback mechanisms can be used to reinforce positive behaviours[20]. These questions are often explored by analysing stylized social dilemmas, such as the public goods game[21] or the collective-risk dilemma[22], that provide valuable insights into the dynamics of cooperation in controlled settings. When subjects interact in such games over multiple rounds, it is typically assumed that the public good remains constant in time, independent of the outcome of previous interactions[9–15]. Here, we explore the emergence of reciprocity when strategic choices in one round affect game payoffs in subsequent rounds. We introduce a framework that allows us to capture the idea that humans affect and are affected by the value of the public resource, and that they are able to anticipate and to adapt to such endogenous changes.

Our approach is based on the theory of stochastic games[16,17]. A group of players can find itself in one of multiple states (Fig. 1). The different states capture how the present physical or social environment affects the feasible actions of the players and their payoffs. The theory of stochastic games[16–19] has applications in computer science[23,24], industrial organization, capital accumulation and resource extraction[17].

We consider stochastic games where, in each state, players interact in a social dilemma with different payoff values. The decision by the players of whether to cooperate or to defect not only affects their current payoffs but also the game that will be played in the next round. In Fig. 1 we illustrate a scenario that reflects the tragedy of the commons. Mutual cooperation improves the quality of the public resource, leading the players to interact in game 1 with comparably high payoffs. Partial defection leads to a deterioration of the resource; players move to game 2 where payoffs are lower. The stochastic game is played for many rounds. Transitions between different states can be stochastic or deterministic, state-dependent or state-independent. The well-studied framework of repeated games is a special case of stochastic games with only one state.

The effect of changing environments on evolutionary dynamics has been explored previously in one-shot, non-repeated games, not using the theory of stochastic games[25–29] (see Supplementary Information, section 1.1). In some scenarios, the co-evolution of the players' strategies and their environment can lead to oscillations between cooperators and defectors[27,28]. But if cooperators are at a disadvantage in every environment, environmental feedback is ineffective to prevent cooperators from going extinct (Supplementary Information). One-shot models assume that players consider only their present payoff when making strategic choices. In stochastic games, players take a long-term perspective instead. To find optimal strategies, they need to consider how their actions affect the response of their opponents and the future state of the environment. As we show, this interplay between reciprocity and payoff feedback can be crucial for cooperation.

Traditionally, work on stochastic games considers rational players who can employ arbitrarily complex strategies, but does not focus on the dynamics of how players adapt their strategies. We introduce an evolutionary perspective to stochastic games. Players do not need to act rationally, but instead they experiment with available strategies and imitate others depending on success[30]. We use simple strategies that are easy to implement and to interpret[8]. Such an evolutionary set-up has proved useful to understand the dynamics of cooperation in repeated games[8–13].

We first study a stochastic game with two states (Fig. 2). Individuals use pure 'memory one' strategies whereby a player's move depends on only the present state and the outcome of the previous round (see Methods and Supplementary Information for details). We compare the stochastic game with the two associated repeated games where the same game occurs every round (Fig. 2). We consider two-player interactions that represent prisoner's dilemmas, as well as n-player public-goods games. In both cases, cooperation entails a cost $c > 0$. In the prisoner's dilemma, cooperation yields a benefit $b_i > c$ to the co-player, where b_i depends on the state i. In the public goods game, aggregated costs are multiplied by a factor r_i (with $1 < r_i < n$ depending on state i), and redistributed among all players. Game 1 is more profitable than game 2 if $b_1 > b_2$ or $r_1 > r_2$. Players find themselves in game 1 only if everyone has cooperated in the previous round. Our simulations show that this feedback can boost cooperation markedly. For reasonable parameters, the stochastic game populations adapt quickly towards full cooperation, although neither of the two repeated games alone yields substantial cooperation levels.

In the stochastic game, cooperation evolves because defectors lose out twice: once, because they risk receiving less cooperation from reciprocal co-players in future and second, because players collectively move towards a less beneficial game. The stochastic game is most effective in boosting cooperation if the benefit in game 1 is intermediate (Extended Data Fig. 1). If b_1 is too low, the double loss present in the stochastic game is not sufficient to incentivize mutual cooperation, whereas if b_1 is high, players cooperate in the first game anyway. Stochastic games can lead to cooperation even if all individual repeated games fail.

[1]Program for Evolutionary Dynamics, Harvard University, Cambridge, MA, USA. [2]IST Austria, Klosterneuburg, Austria. [3]Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic. [4]Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge, MA, USA. *e-mail: christian.hilbe@ist.ac.at; krishnendu.chatterjee@ist.ac.at; martin_nowak@harvard.edu

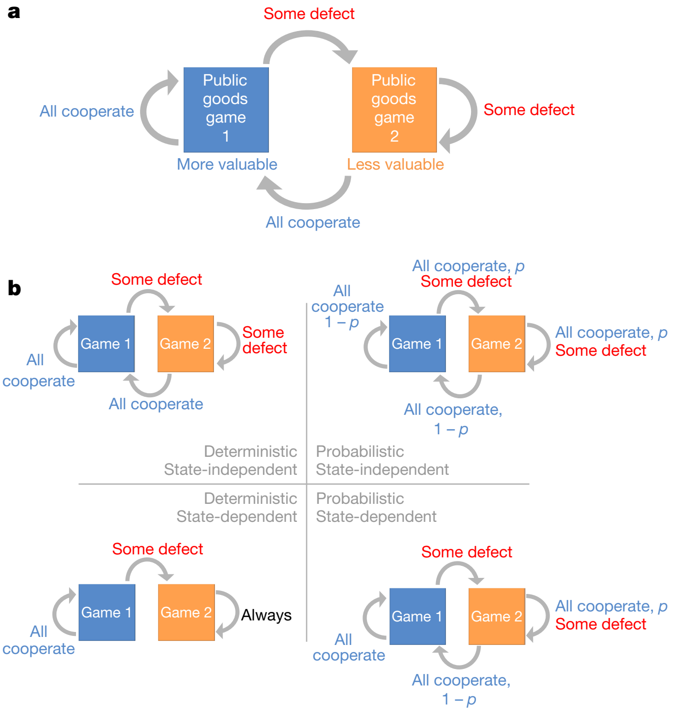


Fig. 1 | In stochastic games, the decisions made by players in one round determine the game that will be played next round. a, For example, if some players defect in a public-goods game, then the environment could deteriorate and thereby reduce the value of the public good. If all cooperate, then the environment could recover and the original value of the public good might be restored. The different states of the environment correspond to the different games that can be played. In this illustration, we show two public-goods games with $r_1 > r_2$. **b**, A stochastic game is deterministic if the players' actions and the current game uniquely determine the game that will be played next round. It is state-independent if the game in the next round depends on only the players' actions, not the current game (state). Thus, we distinguish four different types of stochastic game, depending on whether transitions are deterministic or probabilistic (where p and $1 - p$ indicate the probability of making the respective transition), and whether they are state-independent or state-dependent. We note that even a game that involves only deterministic transitions is referred to as a 'stochastic' game, because it represents a special case of the framework.

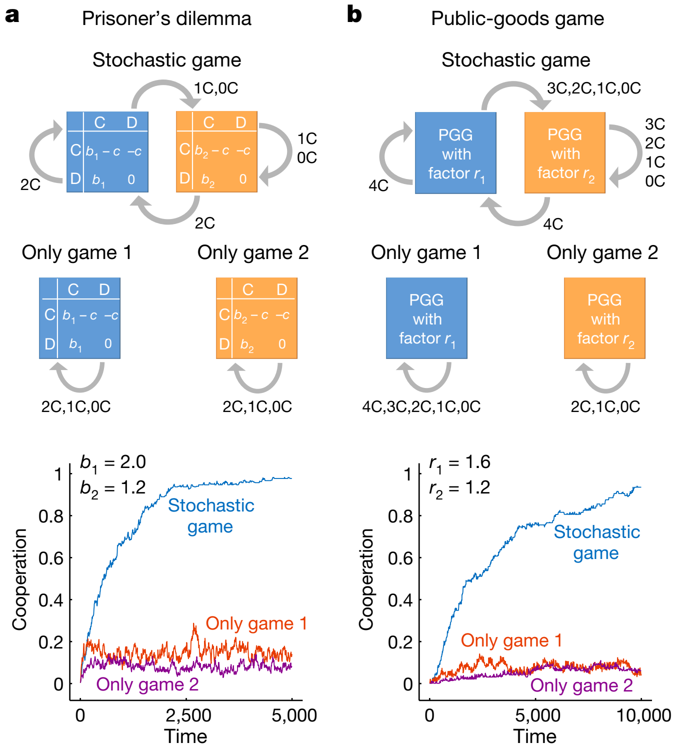


Fig. 2 | Stochastic games can promote cooperation even if all individual games favour defection. a, b, We study the repeated prisoner's dilemma, which is a two-player game (**a**), and the repeated public-goods game (PGG), which is interpreted here as a four-player game (**b**). In both cases, the first game has a higher benefit from cooperation than the second game. Arrows represent the possible transitions, and the arrow labels indicate the number of co-operators ('C') required for the respective transition. The two-player games are represented by their payoff matrices. In the stochastic game, if all players cooperate then the next round will be the first game, but if some players defect ('D') then the next round will be the second game. In the standard repeated games, the same game is used in every round. An analysis based on evolutionary dynamics reveals that each of the standard repeated games fails to support cooperation, whereas the stochastic game favours cooperation. The time axis corresponds to the number of mutant strategies introduced into the population (see Methods). Parameter values: **a**, $b_1 = 2$, $b_2 = 1.2$, $c = 1$; **b**, $r_1 = 1.6$, $r_2 = 1.2$, $c = 1$.

We derive a condition for the stability of cooperation in stochastic games with two states and state-independent transitions. A numerical analysis for the two-player case suggests that full cooperation emerges when win-stay lose-shift[9] (WSLS) becomes stable (Extended Data Figs. 2, 3). This strategy prescribes cooperation in the next round if and only if both players used the same action in the previous round. In a conventional repeated prisoner's dilemma, WSLS is a Nash equilibrium if $b \geq 2c$ (ref. [8]). In the stochastic game, WSLS is an equilibrium if

$$(2q_2 - q_0)b_1 + (1 - 2q_2 + q_0)b_2 \geq 2c \qquad (1)$$

where the parameters q_i refer to the conditional probability that the players will be in game 1 in the next round given that i of them have cooperated in the present round. If mutual cooperation leads to game 1 and mutual defection to game 2, then $q_2 = 1$ and $q_0 = 0$. Therefore, WSLS is stable if $2b_1 - b_2 \geq 2c$. Because $b_1 > b_2$, this condition is easier to satisfy than the respective conditions for the two associated repeated games.

The condition in equation (1) highlights the fact the that the stability of cooperation depends on how the states change given the players' decisions. To explore the effect of this exogenous feedback systematically, we perform simulations for all eight deterministic and state-independent two-state games (Extended Data Fig. 2). In six of the eight cases, players spend more time in the profitable game 1. But only in two of them do players actually cooperate. In line with equation (1), cooperation evolves only if $q_2 = 1$ and $q_0 = 0$, with q_1 being irrelevant. Stochastic games are most effective in promoting cooperation if mutual cooperation improves the public good while mutual defection deteriorates it—a natural scenario. Analogous conclusions hold for multiplayer interactions (Extended Data Figs. 4, 5).

Probabilistic transitions can further enhance the evolution of cooperation. In Fig. 3a, mutual cooperation in game 2 leads back to game 1 with probability q. The optimal value of q is intermediate: players should have some chance to return to the better state, but it should not be too easy (see also Extended Data Fig. 6). In Fig. 3b, the length of the game is not exogenously given, but affected by the players' decisions. Individuals start in state 1, in which they play a conventional prisoner's dilemma; if one or both players defect, then there is some probability q that players move towards state 2, in which no further profitable interactions are possible. This form of environmental feedback promotes cooperation; payoffs become maximal for small but positive q (Extended Data Fig. 7). In Fig. 3c we consider a model with timeout. Defection leads to a temporal state in which no profitable interactions are possible. The return probability to the regular game is q. We derive adaptive dynamics for simple reactive strategies (x, y), where x denotes the cooperation probability after having been in state 1 previously and y is the cooperation probability after having been in timeout. We find

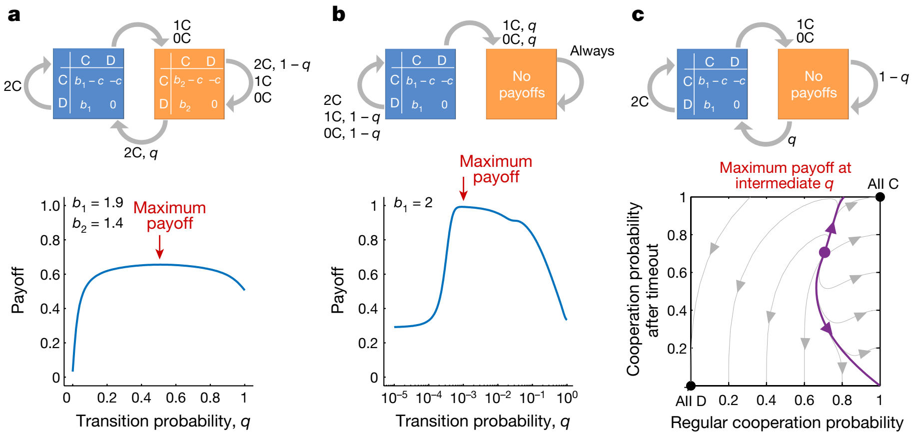


Fig. 3 | Probabilistic transitions maximize cooperation in three different stochastic games. a, Game 1 is more profitable than game 2, but mutual cooperation in game 2 leads to game 1 only with probability q. The evolving average payoffs are maximized for intermediate q. **b**, Game 1 is left with probability q if at least one player has defected. The optimal value of q is small but positive for games with a finite number of rounds (continuation probability $\delta = 0.999$). **c**, Defection leads to a timeout with an expected duration that depends on the return probability q. We derive the adaptive dynamics for strategies that take into account only whether players have been in game 1 in the previous round or in the timeout. Depending on the parameters, 'All C' is a stable endpoint of evolution because no nearby mutant strategy can yield a higher payoff. Again the optimal value of q is intermediate: low values of q increase the area of the phase space for which populations move towards cooperation, but they also make occasional errors more costly (parameters $b_1 = 3$, $c = 1$, $q = 1/2$).

that the fully cooperative strategy (1, 1) can become stable, although unconditional cooperation is never stable in a conventional repeated prisoner's dilemma.

Next we explore the ideal feedback between game payoff and strategic choice. We consider a stochastic game with four players and five states. Defection by a subgroup of players has an immediate, gradual or delayed negative impact on the benefits of cooperation, or no effect (Fig. 4). We obtain the highest cooperation rates for immediate negative impact. The intuitive explanation is as follows: maximum cooperation arises if the players are most incentivized to cooperate in the most valuable game. In the immediate scenario, any deviation from cooperation in game 1 leads to a game with the lowest payoff. Interestingly, even the scenario with a delayed response promotes higher cooperation rates than the game in which the public good remains unchanged across all states. The lowest cooperation rates are obtained when the benefits of cooperation are high in all five games. We obtain similar conclusions for a state-dependent game in which it takes several successive rounds of mutual defection to end up in the worst state (Extended Data Figs. 8, 9).

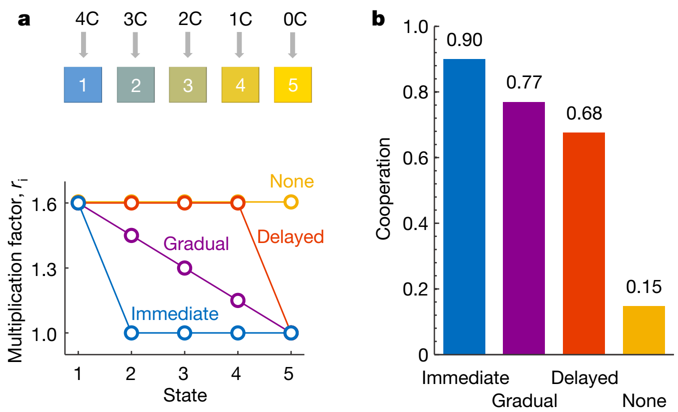


Fig. 4 | Strong immediate feedback maximizes cooperation. a, A four-player scenario in which cooperation improves and defection reduces the value of the public good. Transitions are state-independent: the next state depends on only the number of co-operators, not the previous state. In game 1, contributions to a public good are multiplied by the highest factor $r_1 = 1.6$. In game 5, cooperation does not produce any social benefit, $r_5 = c = 1$. For the payoff in the intermediate games 2, 3 and 4, we distinguish three cases: partial defection has immediate, gradual or delayed consequences on the multiplication factor of the public good. In addition, we consider a fourth scenario in which the multiplication factor remains high in all states ('none', no payoff consequences). **b**, An evolutionary analysis confirms that immediately deteriorating public resources are most favourable to cooperation because they make unilateral exploitation a risky strategy. However, all three stochastic games in which the benefits of cooperation vary lead to substantially more cooperation than the game with no environmental feedback.

Direct reciprocity is a mechanism for the evolution of cooperation based on repeated interactions. The standard assumption has been that the same game, with the same payoff, is played again and again. We have introduced the concept that the game payoff changes in different rounds. We explore cases in which cooperation leads to a more valuable game next round and defection to a less valuable one. Surprisingly, we find that this setting boosts cooperation markedly. In the resulting stochastic game, cooperation can prevail even if it is unsuccessful in all individual repeated games. Our observations suggest how naturally occurring or designed feedback can promote cooperation. A tragedy of the commons can be avoided if the environment deteriorates (rapidly) as a consequence of defection. Likewise, cooperation is boosted if there is the prospect of playing for higher gains should the current cooperation succeed. The evolutionary analysis of stochastic games represents a new tool for understanding and influencing human decision-making in social dilemmas.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0277-x.

Received: 1 November 2017; Accepted: 17 May 2018;
Published online: 04 July 2018

1. Lloyd, W. F. *Two Lectures on the Checks to Population* (Oxford Univ. Press, Oxford, 1833).
2. Hardin, G. The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
3. Trivers, R. L. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
4. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, NY, 1984).
5. Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge Univ. Press, Cambridge, 1990).
6. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
7. Van Lange, P. A. M., Balliet, D., Parks, C. D. & Van Vugt, M. *Social Dilemmas – The Psychology of Human Cooperation* (Oxford Univ. Press, Oxford, 2015).

8. Sigmund, K. *The Calculus of Selfishness* (Princeton Univ. Press, Princeton, 2010).
9. Nowak, M. & 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).
10. Hauert, C. & Schuster, H. G. Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. *Proc. R. Soc. Lond. B* **264**, 513–519 (1997).
11. Killingback, T. & Doebeli, M. The continuous prisoner's dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *Am. Nat.* **160**, 421–438 (2002).
12. Szolnoki, A., Perc, M. & Szabó, G. Phase diagrams for three-strategy evolutionary prisoner's dilemma games on regular graphs. *Phys. Rev. E* **80**, 056104 (2009).
13. Grujić, J., Cuesta, J. A. & Sánchez, A. On the coexistence of cooperators,defectors and conditional cooperators in the multiplayer iterated Prisoner's Dilemma. *J. Theor. Biol*. **300**, 299–308 (2012).
14. García, J. & van Veelen, M. In and out of equilibrium I: evolution of strategies in repeated games with discounting. *J. Econ. Theory* **161**, 161–189 (2016).
15. Hilbe, C., Chatterjee, K. & Nowak, M. A. Partners and rivals in direct reciprocity. *Nat. Hum. Behav*. (2018).
16. Shapley, L. S. Stochastic games. *Proc. Natl Acad. Sci. USA* **39**, 1095–1100 (1953).
17. Neyman, A. & Sorin, S. (eds) *Stochastic Games and Applications* (Kluwer Academic Press, Dordrecht, 2003).
18. Mertens, J. F. & Neyman, A. Stochastic games. *Int. J. Game Theory* **10**, 53–66 (1981).
19. Mertens, J. F. & Neyman, A. Stochastic games have a value. *Proc. Natl Acad. Sci. USA* **79**, 2145–2146 (1982).
20. Rand, D. G. & Nowak, M. A. Human cooperation. *Trends Cogn. Sci*. **17**, 413–425 (2013).
21. Ledyard, J. O. in *The Handbook of Experimental Economics* (eds Kagel, J. H. & Roth, A. E.) 111–194 (Princeton Univ. Press, Princeton, 1995).
22. Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A. & Marotzke, J. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc. Natl Acad. Sci. USA* **105**, 2291–2294 (2008).
23. Alur, R., Henzinger, T. & Kupferman, O. Alternating-time temporal logic. *J. Assoc. Comput. Mach*. **49**, 672–713 (2002).
24. Miltersen, P. B. & Sorensen, T. B. A near-optimal strategy for a heads-up no-limit texas hold'em poker tournament. In *Proc. 6th International Joint Conference on Autonomous Agents and Multiagent Systems* 191 (ACM, 2007).
25. Ashcroft, P., Altrock, P. M. & Galla, T. Fixation in finite populations evolving in fluctuating environments. *J. R. Soc. Interface* **11**, 20140663 (2014).
26. Gokhale, C. S. & Hauert, C. Eco-evolutionary dynamics of social dilemmas. *Theor. Popul. Biol*. **111**, 28–42 (2016).
27. Hauert, C., Holmes, M. & Doebeli, M. Evolutionary games and population dynamics: maintenance of cooperation in public goods games. *Proc. R. Soc. Lond. B* **273**, 2565–2570 (2006); corrigendum 273, 3131–313 (2006).
28. Weitz, J. S., Eksin, C., Paarporn, K., Brown, S. P. & Ratcliff, W. C. An oscillating tragedy of the commons in replicator dynamics with game-environment feedback. *Proc. Natl Acad. Sci. USA* **113**, E7518–E7525 (2016).
29. Tavoni, A., Schlüter, M. & Levin, S. The survival of the conformist: social pressure and renewable resource management. *J. Theor. Biol*. **299**, 152–161 (2012).
30. Traulsen, A., Nowak, M. A. & Pacheco, J. M. Stochastic dynamics of invasion and fixation. *Phys. Rev. E* **74**, 011909 (2006).

Acknowledgements 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 support from the ISTFELLOW programme.

Reviewer information *Nature* thanks A. Neyman and the other anonymous reviewer(s) for their contribution to the peer review of this work.

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

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0277-x.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0277-x.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.H. or K.C. or M.A.N.

METHODS

Here we summarize our general framework and the methods that we used. Further details are provided in Supplementary Information.

Stochastic games. To describe a stochastic game fully, we need to specify five objects: (i) the set of players $\mathcal{N}$, (ii) the set of possible states S, (iii) the set of actions $A(s_i)$ that are available to each player in a given state s_i, (iv) the transition function Q that describes how the current state of the environment and the players' actions in a given round determine the state in the next round, and (v) a payoff function u that describes how the payoffs of the players in a given round depend on the players' actions and on the present state. The framework of stochastic games does not specify how much time passes between consecutive rounds, nor does it restrict the payoffs that are available in each round. The respective model parameters need to be chosen with respect to the specific application (see Supplementary Information for a detailed description of the framework and how it applies to specific examples). Here we have considered scenarios in which players face a strict social dilemma in each state, but the framework can easily be adapted to more general payoff constellations (Extended Data Fig. 10).

Throughout the main text, we considered simple examples of stochastic games. Players can choose between cooperation and defection, and thus their action set is {C, D} for each state. Transitions are symmetric: the transition function Q does not depend on which of the players has cooperated or defected. The payoffs per round are symmetric and in the two-player case given by payoff matrices. The payoff of a player in the stochastic game is defined as the player's discounted payoff per round over infinitely many rounds. Initially, players are in state 1. Here we focus on stochastic games that take place in discrete time, but continuous-time stochastic games have also been considered[31] (see Supplementary Information for a more detailed discussion).

Memory-one strategies. In general, strategies for stochastic games can be arbitrarily complex. A player's action in a given round may depend on the present state and on the whole previous history. To facilitate an evolutionary analysis, we focus on comparably simple strategies[32–39]: players take into account only the present state and the outcome of the previous round. For n-player games with m states, such 'memory one' strategies can be written as a $2nm$-dimensional vector $\boldsymbol{p} = (p^i_{a,j})$, with $i \in \{1, 2, \ldots, m\}$, $j \in \{0, 1, \ldots, n-1\}$ and $a \in \{\mathrm{C}, \mathrm{D}\}$. Each entry $p^i_{a,j}$ represents the player's probability of cooperating in a given round, given that the present state is s_i and that in the previous round the focal player chose action $a \in \{\mathrm{C}, \mathrm{D}\}$, while j of the $n-1$ other group members cooperated. In Supplementary Table 1, we present several examples of memory-one strategies for stochastic games.

When all players use memory-one strategies, the dynamics of a stochastic game can be described by a Markov chain with $m2^n$ possible states $(s_1, \mathrm{C}, \ldots, \mathrm{C}), \ldots, (s_m, \mathrm{D}, \ldots, \mathrm{D})$. In this notation, the first entry refers to the state of the public good in a given round and the other n entries refer to the players' actions. Using the theory of Markov chains, we compute the players' expected payoffs (see Supplementary Information).

Evolutionary dynamics. To describe how individuals adopt new strategies over time, we consider a standard imitation process[30]. There is a population of size N. Each member of the population is equipped with a memory-one strategy that prescribes how the individual plays the stochastic game. In each evolutionary time step, every player interacts with every other player to derive a payoff from the stochastic game. Then, two individuals are drawn randomly from the population, a learner and a role model. The payoffs of those two individuals are π_{L} and π_{R}, respectively. The learner adopts the strategy of the role model with probability $\rho = 1/[1+\mathrm{e}^{-\beta(\pi_{\mathrm{R}}-\pi_{\mathrm{L}})}]$. The parameter $\beta \geq 0$ corresponds to the intensity of selection. For $\beta = 0$, we have random drift. For $\beta > 0$, imitation events are biased in favour of strategies that yield higher payoffs. In addition to imitation events, we allow for random strategy exploration, which corresponds to mutations: with probability μ an individual adopts a randomly chosen memory-one strategy instead of imitating a co-player. We analyse the ergodic mutation–selection process using computer simulations. We obtain exact numerical results when exploration events are rare.

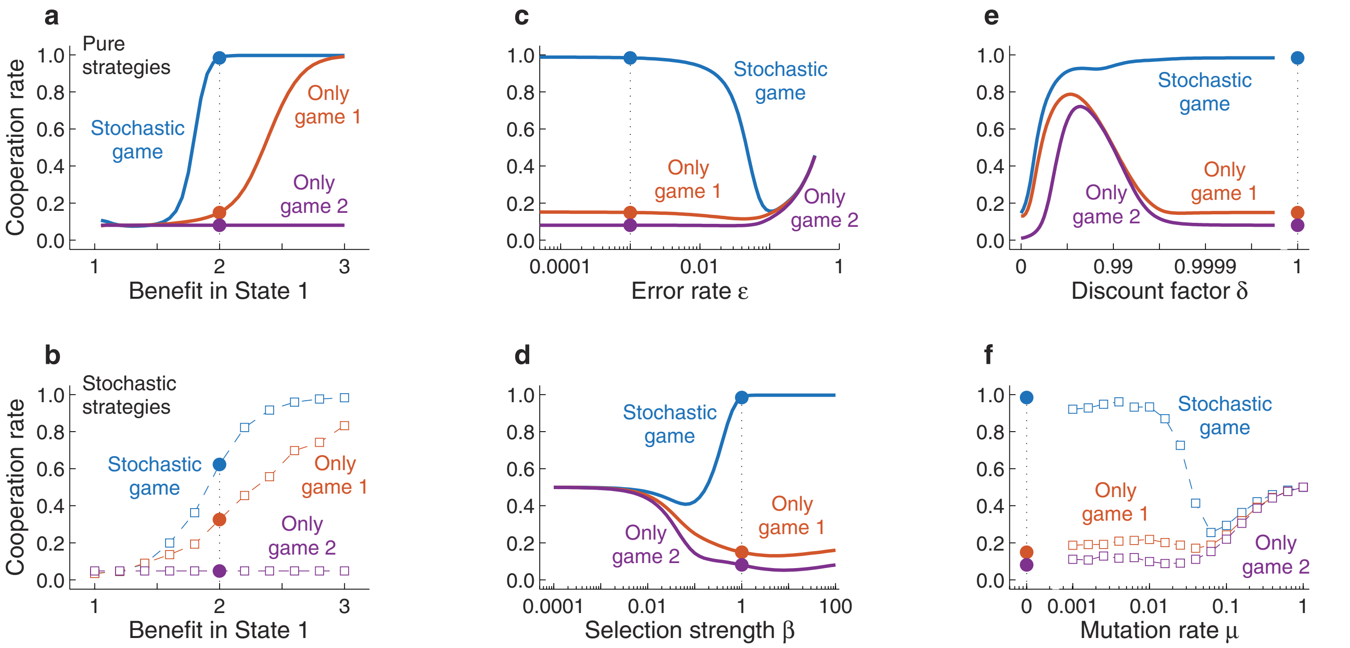
Specific methods used for individual figures. Except for the results in Fig. 3c, the main text considers examples in which players use pure memory-one strategies, subject to small errors (such that $p^i_{a,j}$ is either ε or $1-\varepsilon$, with $\varepsilon = 0.001$). Further simulations using stochastic memory-one strategies confirm that the respective results are robust (Extended Data Fig. 1b). Except for the stochastic game in Fig. 3b, we assume that future payoffs are not discounted, $\delta \to 1$. For the evolutionary trajectories of Fig. 2, we averaged over 100 simulations for the scenario with rare mutations. Our numerical results use population size $N = 100$, intermediate selection ($\beta = 1$) for pairwise games and strong selection for multiplayer games ($\beta = 100$ in Fig. 2b and $\beta = 10$ in Fig. 4). Our qualitative findings are robust with respect to parameter changes (Extended Data Fig. 1). For the results in Fig. 3a, b and 4 we report exact results in the limit of rare mutations[40]. Figure 3c shows the phase portrait of adaptive dynamics[8] for the game with timeout; the corresponding differential equation is derived in Supplementary Information.

Code availability. All simulations and numerical calculations were performed with MATLAB R2014A. In Supplementary Information (see appendix), we provide an algorithm that can be used to calculate payoffs in stochastic games with n players and two states. All other scripts are available from the authors on request or at https://doi.org/10.5281/zenodo.1287718.

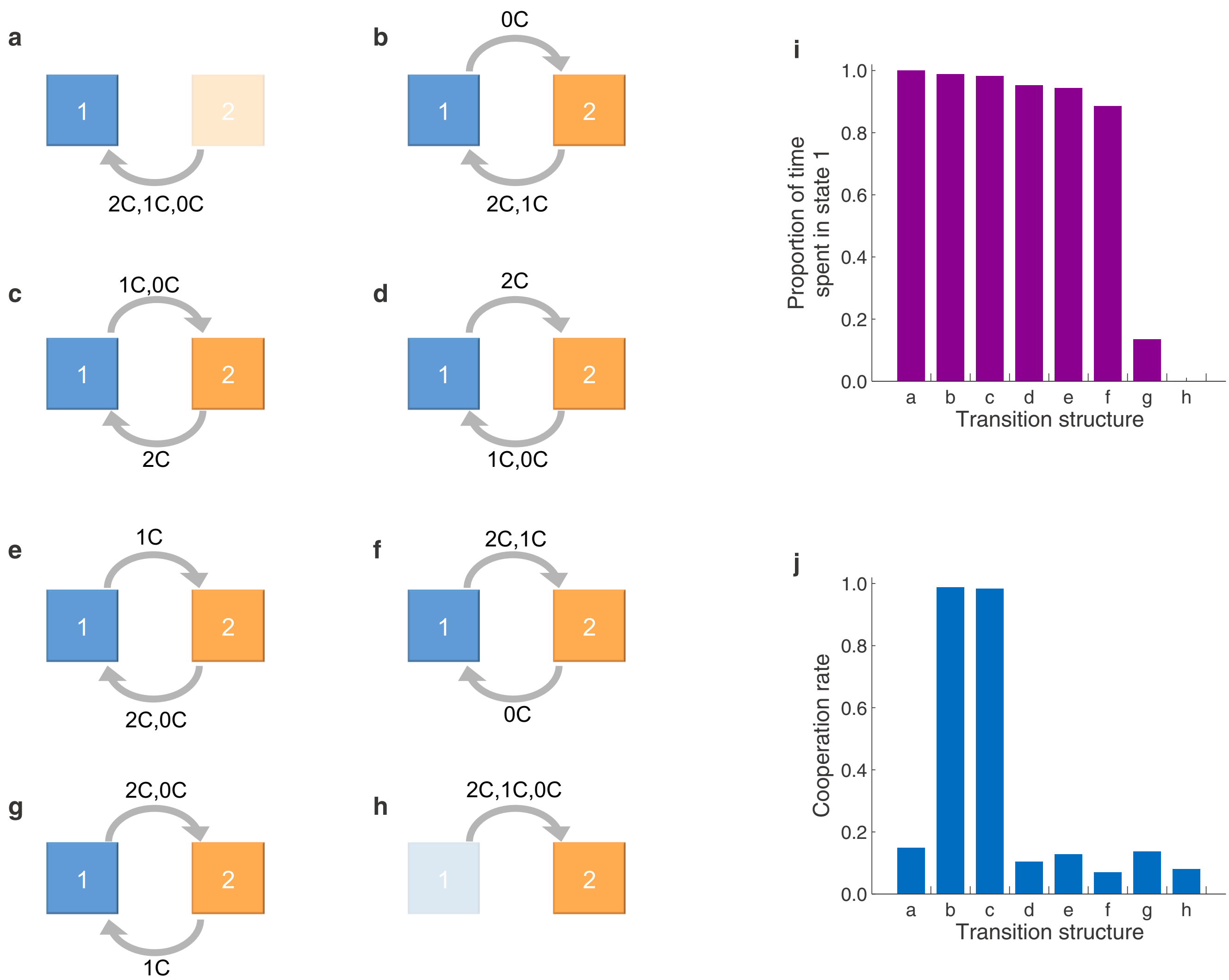
Data availability. The raw data generated, which were used to create Figs. 2–4, have been uploaded along with the MATLAB code and are available at https://doi.org/10.5281/zenodo.1287718.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

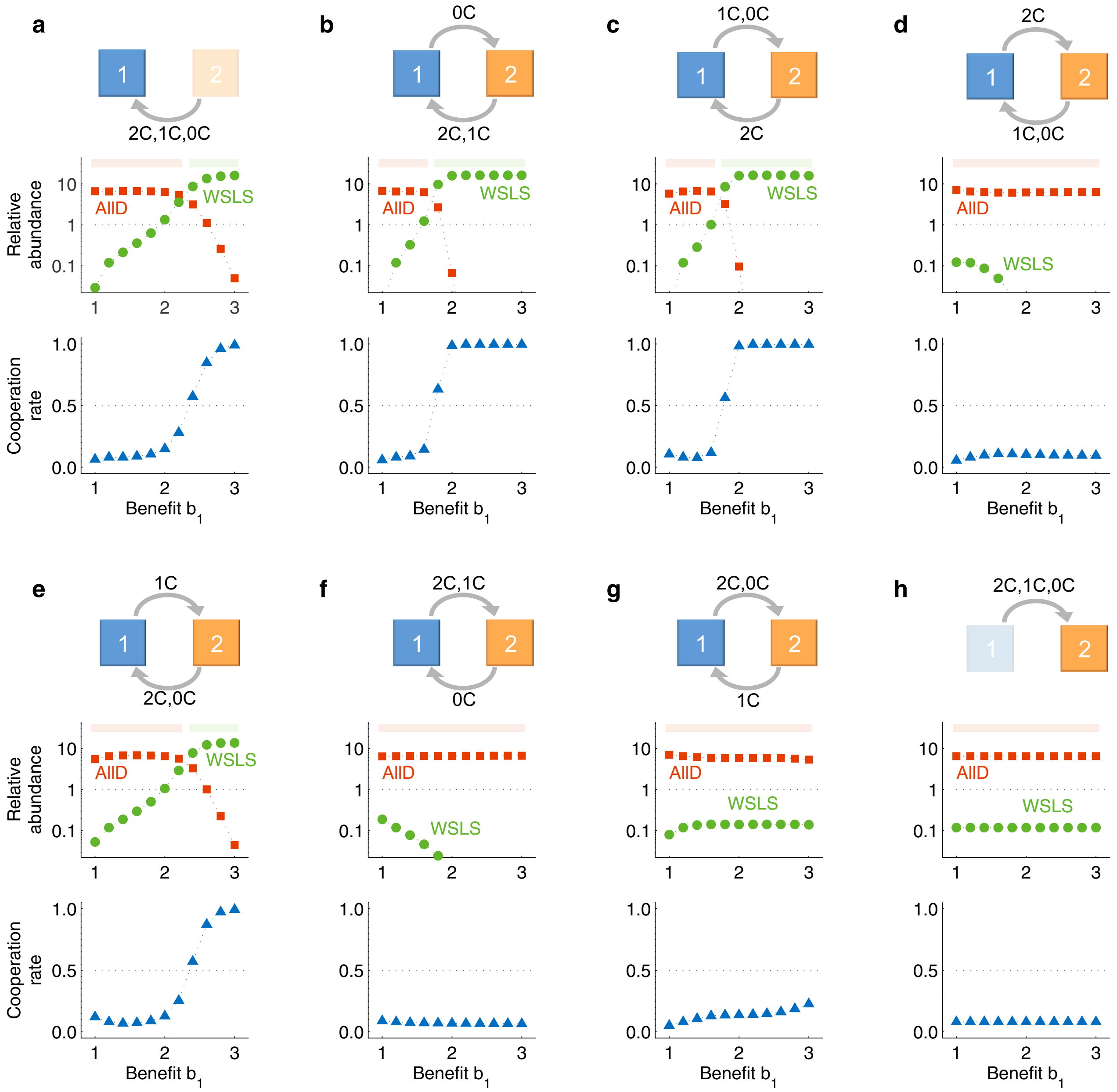
31. Neyman, A. Continuous-time stochastic games. *Games Econ. Behav.* **104**, 92–130 (2017).
32. Nowak, M. A. & Sigmund, K. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Appl. Math.* **20**, 247–265 (1990).
33. Ohtsuki, H. & Iwasa, Y. The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* **239**, 435–444 (2006).
34. Stewart, A. J. & Plotkin, J. B. Collapse of cooperation in evolving games. *Proc. Natl Acad. Sci. USA* **111**, 17558–17563 (2014).
35. 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).
36. Akin, E. in *Ergodic Theory, Advances in Dynamics* (ed. Assani, I.) 77–107 (de Gruyter, Berlin, 2016).
37. Hilbe, C., Martinez-Vaquero, L. A., Chatterjee, K. & Nowak, M. A. Memory-*n* strategies of direct reciprocity. *Proc. Natl Acad. Sci. USA* **114**, 4715–4720 (2017).
38. Stewart, A. J. & Plotkin, J. B. Small groups and long memories promote cooperation. *Sci. Rep.* **6**, 26889 (2016).
39. Reiter, J. G., Hilbe, C., Rand, D. G., Chatterjee, K. & Nowak, M. A. Crosstalk in concurrent repeated games impedes direct reciprocity and requires stronger levels of forgiveness. *Nat. Commun.* **9**, 555 (2018).
40. Fudenberg, D. & Imhof, L. A. Imitation processes with small mutations. *J. Econ. Theory* **131**, 251–262 (2006).



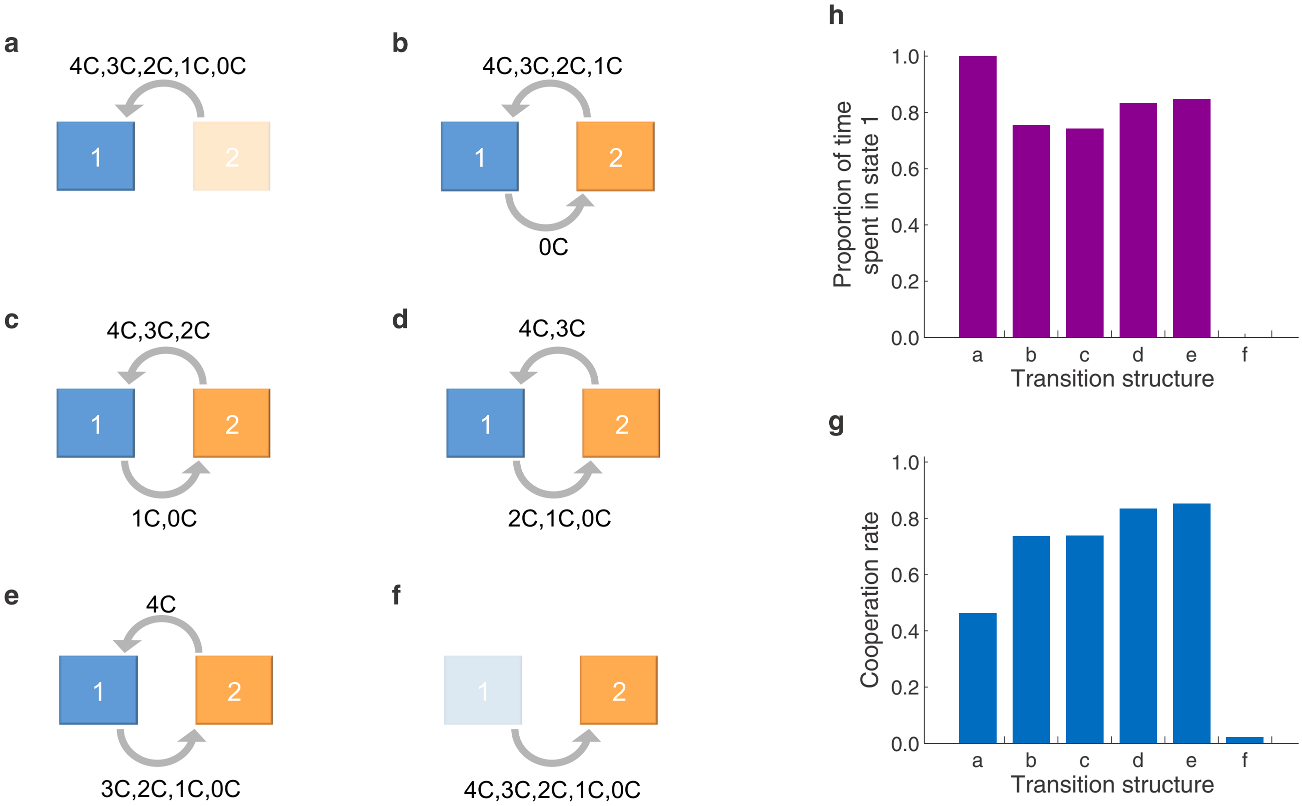
Extended Data Fig. 1 | Our findings are robust with respect to parameter changes. To test the robustness of our findings, we consider the stochastic game introduced in Fig. 2a and independently vary several key parameters. **a**, **b**, When we vary the benefit of cooperation in state 1, we find that the advantage of the stochastic game is most pronounced when this benefit is intermediate, $1.5 \leq b_1 \leq 2.5$. This conclusion holds independently of whether individuals use pure strategies only (**a**) or stochastic ones (**b**). **c**–**f**, We obtain similar results when we vary the error rate ε (**c**), the strength of selection β (**d**), the discount factor δ (**e**) and the mutation rate μ (**f**). In all cases, we observe that stochastic games yield a cooperation premium, provided that errors are sufficiently rare, selection is sufficiently strong, players give sufficient weight to future payoffs and mutations are comparably rare. Solid lines indicate exact results in the limit of rare mutations, whereas square symbols and dashed lines represent simulation results (see Supplementary Information for details). Filled circles highlight the results obtained for the parameters in Fig. 2a. As default parameters, we used the same values as in Fig. 2a: $N = 100$, $b_1 = 2.0$, $b_2 = 1.2$, $c = 1$, $\beta = 1$, $\varepsilon = 0.001$, $\delta \rightarrow 1$ and $\mu \rightarrow 0$.



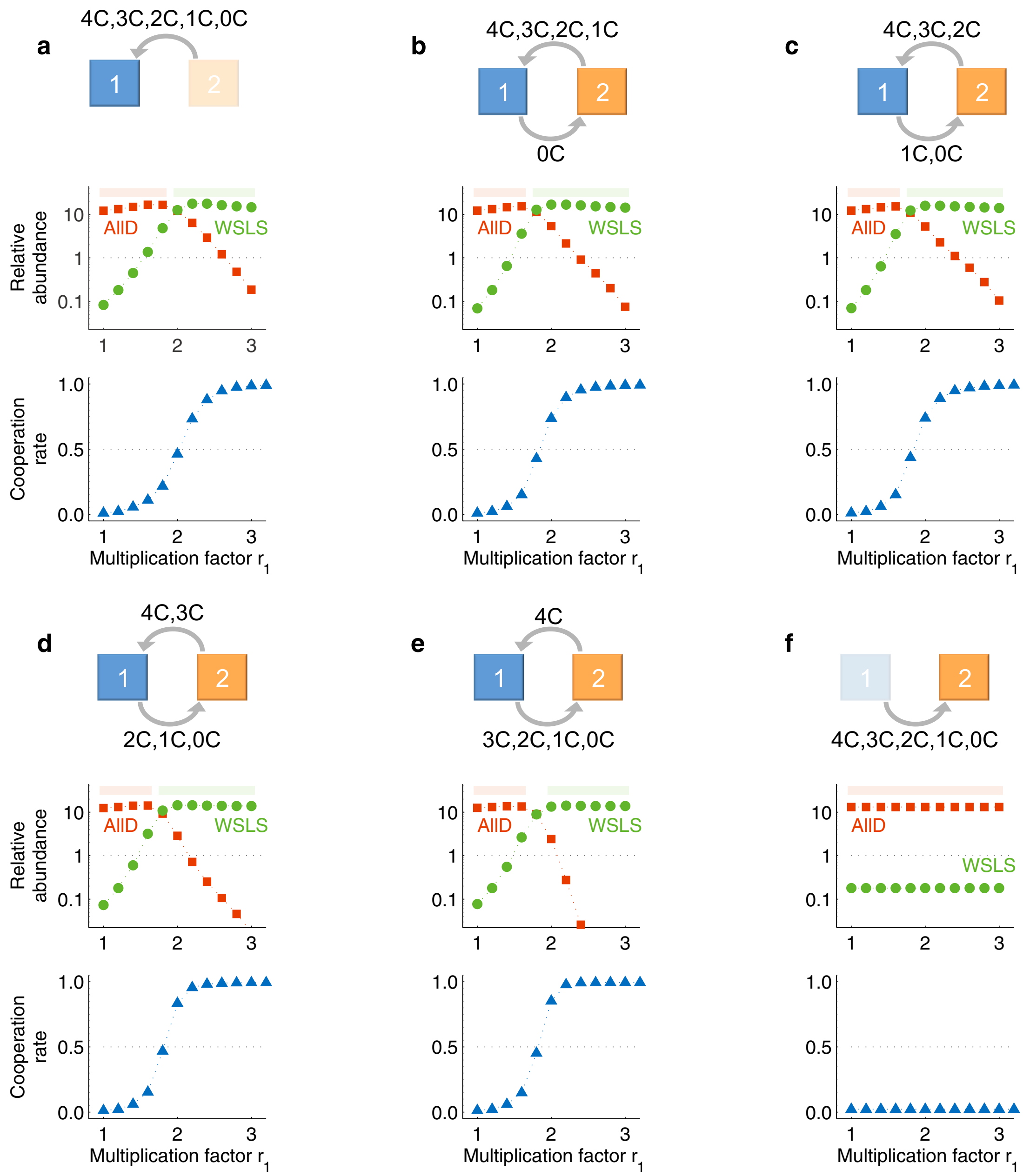
Extended Data Fig. 2 | Whether cooperation evolves in two-player games depends critically on the form of the environmental feedback. Keeping the game parameters fixed at the values used in Fig. 2a, we explored how the evolution of cooperation depends on the underlying transition structure of the stochastic game in the limit of rare mutations (see Supplementary Information). **a–h**, We calculated the selection–mutation equilibrium for all possible stochastic games with two states when transitions are state-independent and deterministic. **i**, Overall, six of the eight transition structures lead players to spend more time in the more profitable state 1, in which mutual cooperation has a higher benefit. **j**, However, cooperation evolves in only two out of these six transition structures. These two structures have in common that mutual cooperation always leads to the beneficial state 1, whereas mutual defection leads to the detrimental state 2. Thus, cooperation is most likely to evolve if the environmental feedback itself incentivizes mutual cooperation and disincentivizes mutual defection. The transitions after unilateral defection have a less prominent role.



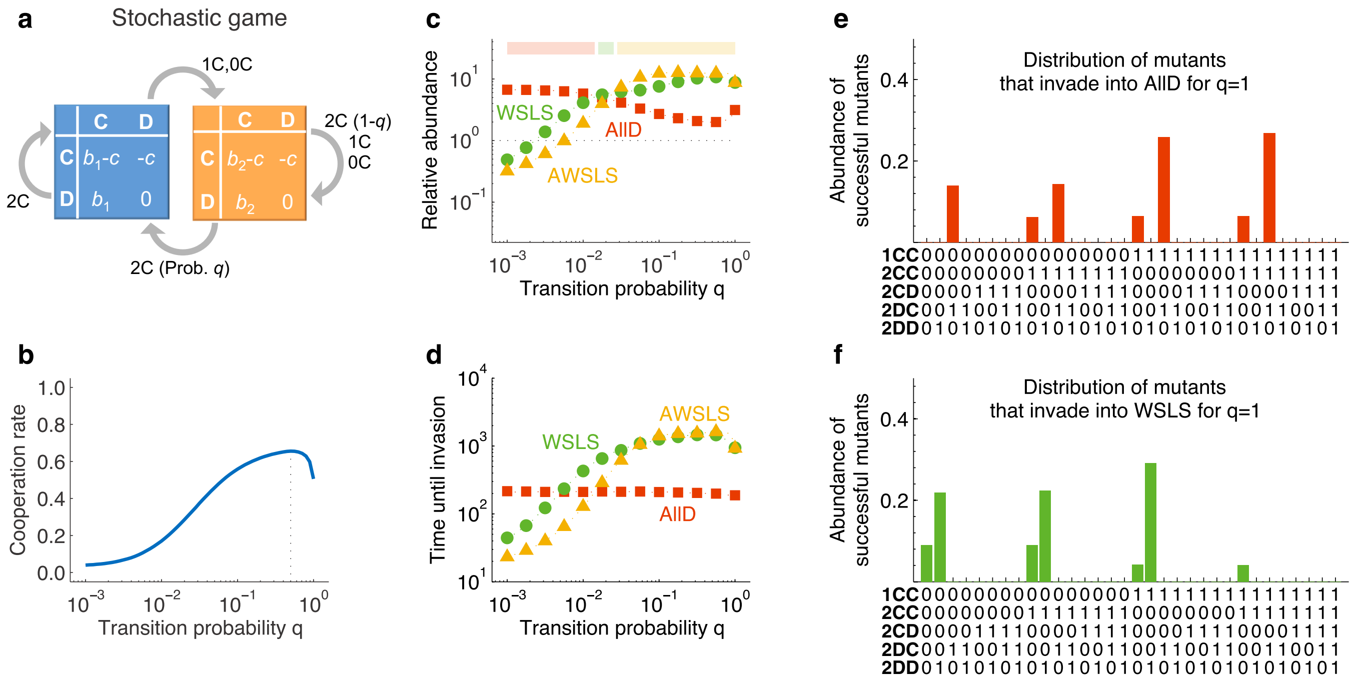
Extended Data Fig. 3 | Analysis of the evolving strategies suggests that the evolution of cooperation hinges on the success of WSLS. Here, we consider all state-invariant and deterministic stochastic games with two states and two players. **a–h**, For each of the eight possible cases, we recorded the evolving cooperation rate (lower plots) and the relative abundance of each pure memory-one strategy (upper plots) for different values of b_1. For clarity, we depict only two memory-one strategies explicitly, All D (the strategy that prescribes to always defect) and WSLS. The colour-shaded bars on top of the upper plots show parameter regimes in which either All D or WSLS is most abundant among all 16 strategies. In four of the eight cases, we observe that full cooperation evolves as the benefit to cooperation in state 1 approaches $b_1 = 3$. These are exactly the cases in which mutual cooperation leads players towards the more beneficial state 1. Moreover, in these four cases the upper plots show that cooperation emerges owing to the success of WSLS, which is the predominant strategy whenever cooperation prevails. Except for the value of b_1, all other parameter values are the same as in Extended Data Fig. 2.



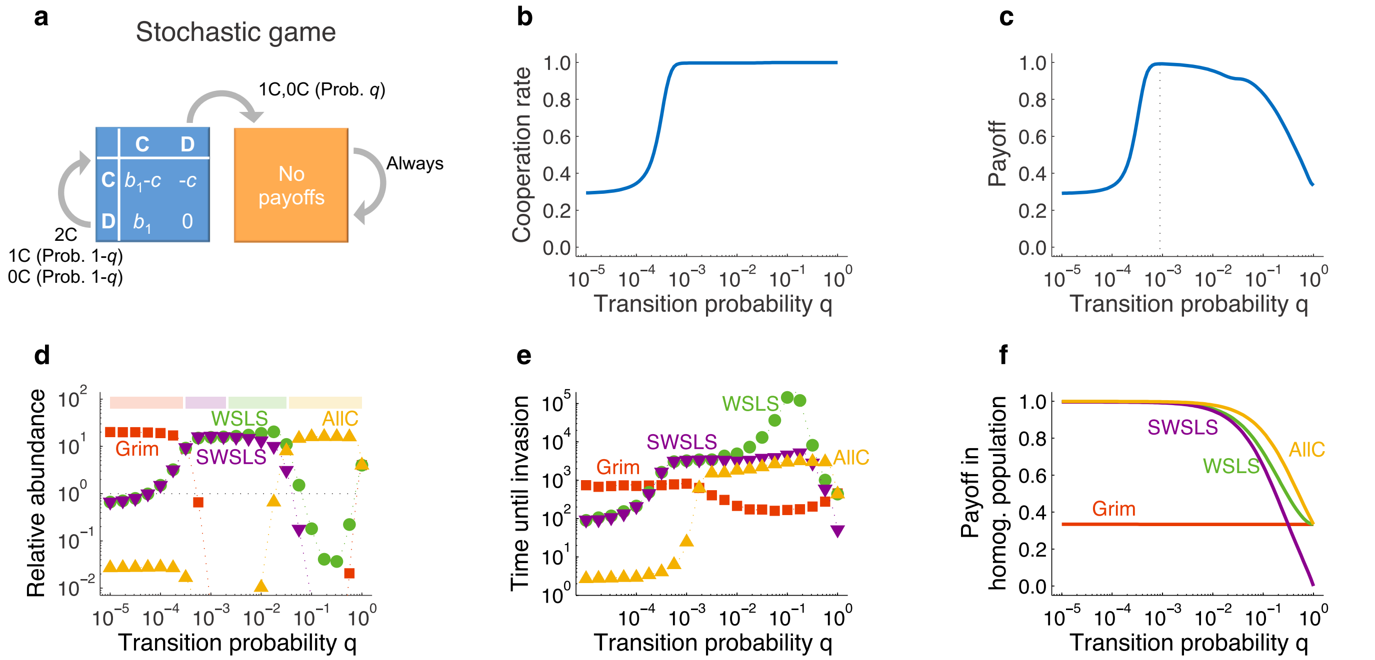
Extended Data Fig. 4 | Effect of transitions on cooperation in four-player public-goods games. We also explored the effect of different transition structures for stochastic games between multiple players (with a public-goods game being played in each state). State 1 is again more beneficial because $r_1 > r_2$, but to be in state 1 there must be a minimum number k of cooperators in the previous round. **a**–**f**, For a four-player public-goods game, there are six possible monotonic configurations of the stochastic game because k can be any number from 0 (players always move to first state) to 5 (players never move to first state). **h**, There is a non-monotonic relationship between the six transition structures and the time spent in the more beneficial state 1. **g**, The evolving cooperation rate becomes maximal when any deviation from mutual cooperation leads players to state 2 (**e**). Parameters are as in Fig. 2b, but with the multiplication factor in the first state fixed to $r_1 = 2$ and selection strength $\beta = 1$; to derive exact results, we considered the limit of rare mutations $\mu \rightarrow 0$ (see Supplementary Information for details).



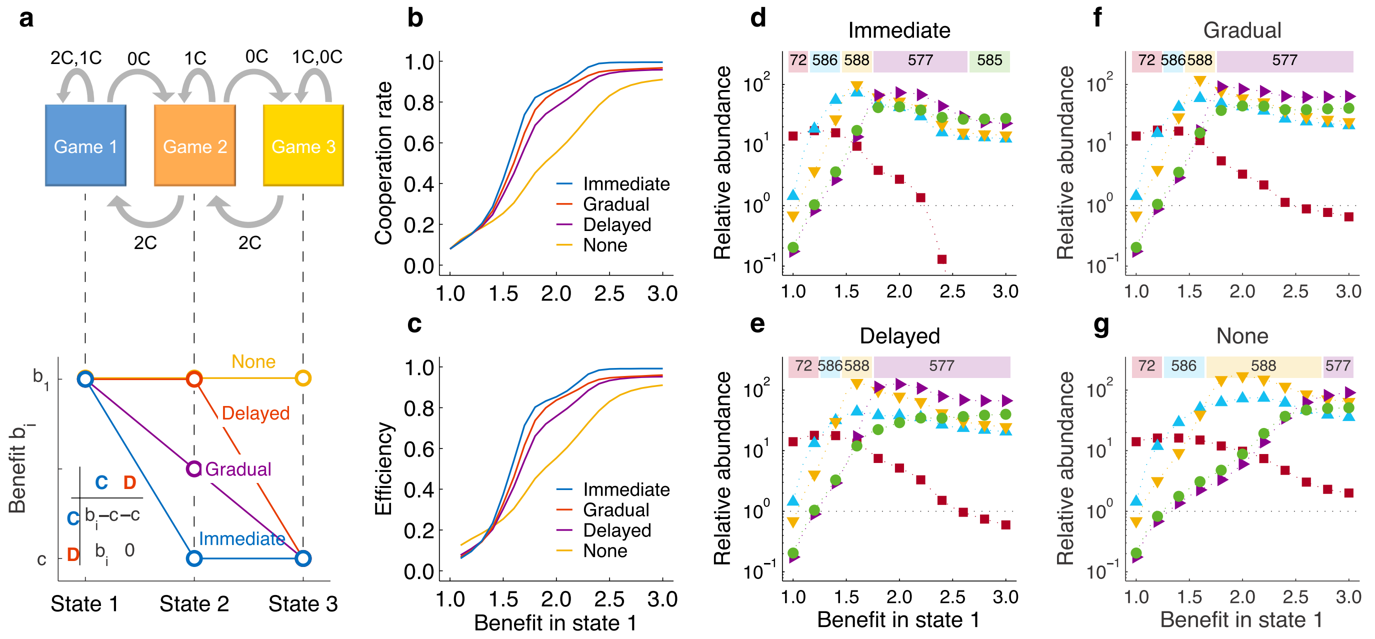
Extended Data Fig. 5 | WSLS sustains cooperation in multiplayer public-goods games. This figure is analogous to Extended Data Fig. 3 for the case of multiplayer interactions. Again, we show evolving cooperation rates and the relative abundance of All D and WSLS for the six state-independent and deterministic games in which transitions are monotonic. In five of these games, cooperation emerges once the multiplication factor r_1 becomes sufficiently large. In all of those, WSLS is the most abundant strategy when cooperation evolves. Except for r_1, all parameters are the same as in Extended Data Fig. 4.



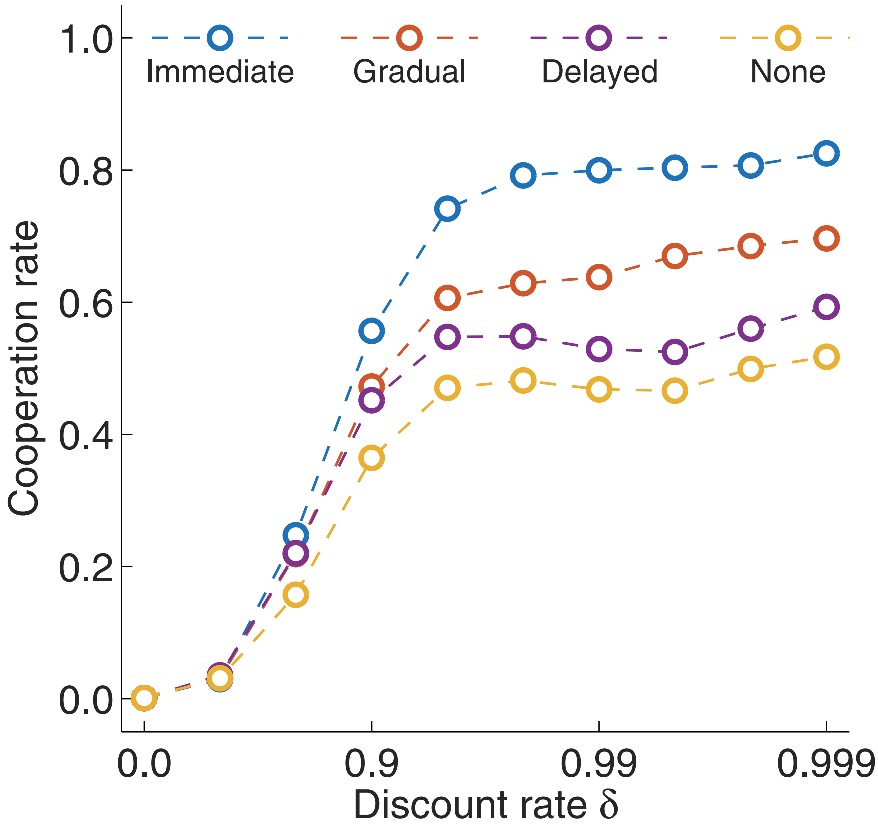
Extended Data Fig. 6 | Probabilistic transitions can further enhance cooperation. a, Here, we explore in more detail the stochastic game introduced in Fig. 3a (see Supplementary Information for details), in which any defection always leads to state 2. After mutual cooperation in state 1, players remain in state 1 with certainty. After mutual cooperation in state 2, players move towards state 1 with probability q. **b**, Calculating the cooperation rate in the selection–mutation equilibrium in the limit of rare mutations shows that the highest cooperation rate is achieved for intermediate values of q. **c**, We recorded the abundance of all 32 memory-one strategies in the selection–mutation equilibrium. The most abundant strategy is either All D (for small values of q, as indicated by the red squares), WSLS (for small but positive values of q, green circles) or AWSLS (for all other values of q, yellow triangles; AWSLS is a more ambitious variant of WSLS, see Supplementary Information, section 4.1). **d**, To estimate the time that it takes each resident strategy to be invaded, we randomly introduced other mutant strategies and recorded how long it took until a mutant successfully fixed (that is, the number of independent mutant strategies introduced before the mutant strategy was adopted by the whole population). To obtain a reliable estimate, we performed 10,000 runs for each resident strategy. **e**, **f**, In addition, we recorded which strategy eventually reaches fixation if the resident applies either All D or WSLS when $q = 1$. Parameters: $b_1 = 1.9$, $b_2 = 1.4$, $c = 1$, $\beta = 1$, $N = 100$.



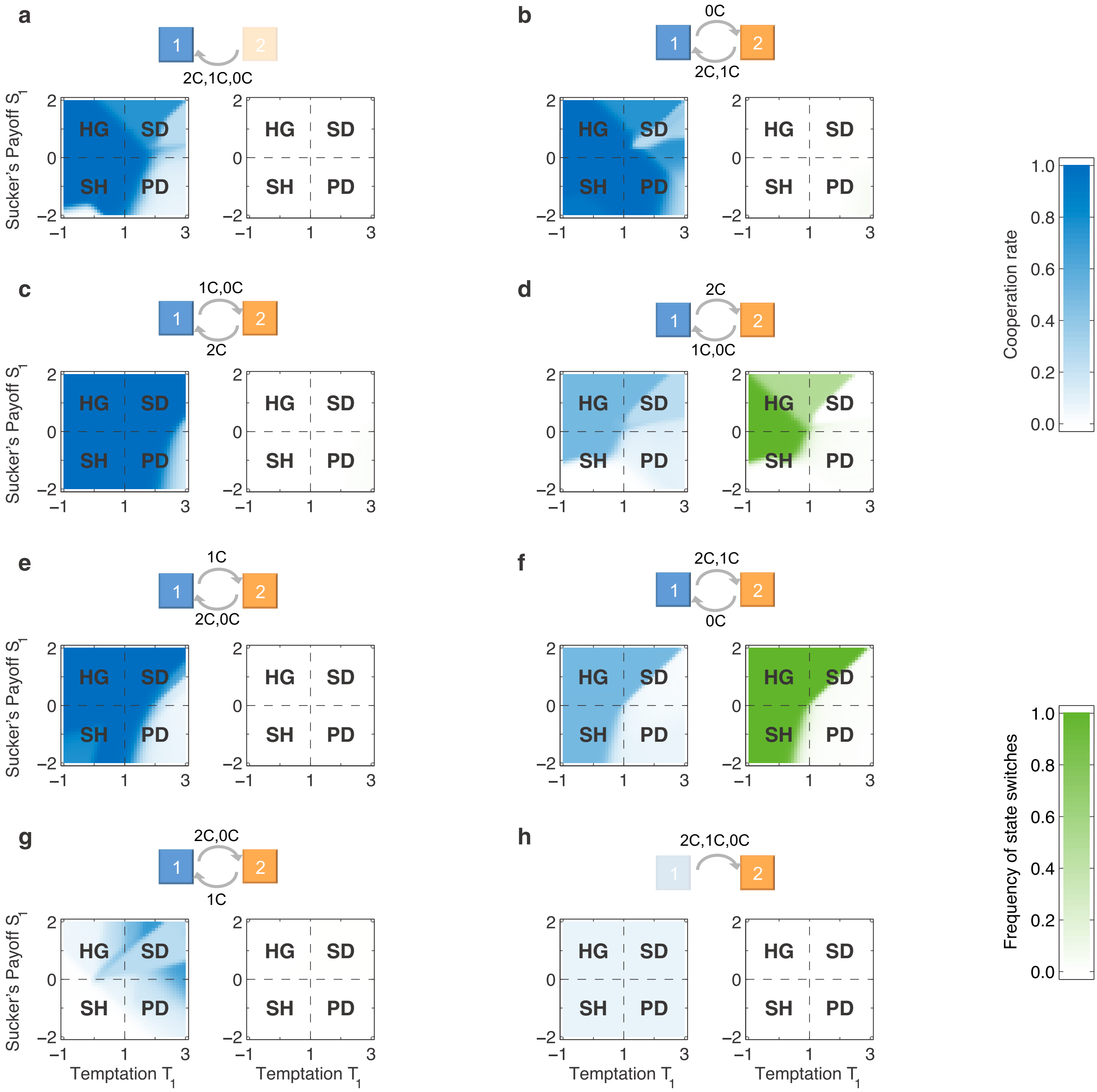
Extended Data Fig. 7 | Players benefit from a small endogenous risk that the game stops early. a, We consider the stochastic game in Fig. 3b, in which players remain in state 1 after cooperation, but move towards state 2 with transition probability q if one of the players defects. In state 2, no profitable interactions are possible. All results are discussed in detail in Supplementary Information; here we provide a summary. **b**, According to our evolutionary simulations, a higher transition probability leads to more cooperation. **c**, However, a higher probability q also makes players move to the second state if one of them defected merely owing to an error; hence, the dependence of payoffs on q is non-monotonic. **d**, **e**, When q is small, Grim is the predominant strategy. Players with this strategy cooperate until one of the players defects; from then on, they defect forever. As q increases, WSLS strategies take over. As $q \to 1$, unconditional cooperation becomes most successful. **f**, For the given parameter values, a homogeneous Grim population achieves only one-third of the maximum payoff possible, because any error leads to relentless defection. The other three strategies result in the maximum payoff $b_1 - c$ for $q = 0$, but this payoff decreases with q. Parameters: $b_1 = 2$, $c = 1$, $\delta = 0.999$, $\varepsilon = 0.001$, $\beta = 1$, $N = 100$.



Extended Data Fig. 8 | Immediate environmental feedback enhances cooperation. a, We consider a state-dependent stochastic game with two players and three states. Mutual cooperation always leads players to move to a superior state (or to remain in the most beneficial state s_1). Similarly, mutual defection always leads to an inferior state (or players remain in the most detrimental state s_3). After a unilateral defection, players remain in the same state. We consider four different versions of this game, depending on how quickly the payoffs decrease as players move towards an inferior state. **b**, Our numerical results show that an immediate negative response of the environment to defection is most favourable to the evolution of cooperation. **c**, As a consequence, the scenario with immediate consequences also yields the highest average payoffs once the benefit in state 1 exceeds a moderate threshold. **d**–**g**, On the level of evolving strategies, we find that an immediately responding environment is most favourable to the evolution of WSLS strategies and strongly selects against defecting strategies. Again, the coloured bars on top of each panel indicate the strategy that is most favoured by selection for the respective value of b_1 (see Supplementary Information for all details). Parameters: $c = 1$; b_1 varies from 1 to 3; b_2 is equal to c, $(b_1 + c)/2$ or b_1; and b_3 is equal to either c or b_1 depending on the scenario considered (as depicted in **a**); $N = 100$, $\beta = 1$, $\delta \to 1$, $\varepsilon = 0.001$.



Extended Data Fig. 9 | Cooperation in stochastic games requires that players take future payoff consequences into account. We repeated the numerical computations in Extended Data Fig. 8 for various discount rates δ. When players focus entirely on the present ($\delta = 0$), cooperation evolves in none of the four treatments. As players increasingly take future payoffs into account, cooperation rates increase. Immediate payoff feedback is most conducive to cooperation across all values of δ considered. Except for the discount rate, parameters are the same as in Extended Data Fig. 8, with $b_1 = 1.8$.



Extended Data Fig. 10 | A systematic analysis of the expected game dynamics for different game payoffs. Keeping the two-player game in state 2 fixed to the game in Fig. 2a, we varied the game that is played in state 1. We assume that payoffs in the first state are 1 (for mutual cooperation), S_1 (for unilateral cooperation), T_1 (for unilateral defection) and 0 (for mutual defection). Depending on T_1 and S_1, game 1 can be one of four different types: harmony game (HG), snowdrift game (SD), stag-hunt game (SH) or prisoner's dilemma (PD); see Supplementary Information for details. For each of the eight possible state-independent transitions $\boldsymbol{q}$, we systematically varied the temptation payoff T_1 (x axis) and the sucker's payoff S_1 (y axis) in the first state (see Supplementary Information for details). For each combination of T_1, S_1 and $\boldsymbol{q}$, we computed how often players cooperate in the selection–mutation equilibrium (left panels) and in what fraction of rounds they switch from one state to the other (right panels). **a–c**, **e**, Full cooperation can evolve when players find themselves in state 1 after mutual cooperation. **d**, **f**, Players learn to switch between states only when mutual cooperation leads to state 2 and mutual defection leads to state 1. **g**, **h**, In the remaining cases, players hardly cooperate. The payoffs in game 2 are the same as in Fig. 2a—a prisoner's dilemma with $b_2 = 1.2$ and $c = 1$. For the evolutionary parameters we considered population size $N = 100$ and selection strength $\beta = 1$.

nature research

Corresponding author(s): Martin A. Nowak

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☒	☐	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	All computations and simulations were performed with with Matlab R2014a. The baseline code is provided in the SI.
Data analysis	Results were analyzed and visualized with Matlab R2014a.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The raw data for Figs. 1-4 generated by the MATLAB programs have been uploaded along with the MATLAB scripts, see Code Availability and Data Availability statements in the end of the Methods section.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Theoretical study that employs analytical methods and evolutionary simulations.
Research sample	n/a (the manuscript does not contain any empirical data)
Sampling strategy	n/a (see above)
Data collection	n/a (see above)
Timing and spatial scale	n/a (see above)
Data exclusions	n/a (no data of any sort was excluded)
Reproducibility	n/a
Randomization	n/a
Blinding	n/a

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Unique biological materials
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☒	☐	Human research participants

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging