Cooperation under Indirect Reciprocity and Imitative Trust

Serguei Saavedra[1,2]*, David Smith[3,4,5], Felix Reed-Tsochas[3,6]

1 Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois, United States of America, 2 Kellogg School of Management, Northwestern University, Evanston, Illinois, United States of America, 3 CABDyN Complexity Centre, Oxford University, Oxford, United Kingdom, 4 Oxford Centre for Integrated Systems Biology, Oxford University, Oxford, United Kingdom, 5 Centre for Mathematical Biology, Oxford University, Oxford, United Kingdom, 6 Institute for Science, Innovation, and Society, Saïd Business School, Oxford University, Oxford, United Kingdom

Abstract

Indirect reciprocity, a key concept in behavioral experiments and evolutionary game theory, provides a mechanism that allows reciprocal altruism to emerge in a population of self-regarding individuals even when repeated interactions between pairs of actors are unlikely. Recent empirical evidence show that humans typically follow complex assessment strategies involving both reciprocity and social imitation when making cooperative decisions. However, currently, we have no systematic understanding of how imitation, a mechanism that may also generate negative effects via a process of cumulative advantage, affects cooperation when repeated interactions are unlikely or information about a recipient's reputation is unavailable. Here we extend existing evolutionary models, which use an image score for reputation to track how individuals cooperate by contributing resources, by introducing a new imitative-trust score, which tracks whether actors have been the recipients of cooperation in the past. We show that imitative trust can co-exist with indirect reciprocity mechanisms up to a threshold and then cooperation reverses -revealing the elusive nature of cooperation. Moreover, we find that when information about a recipient's reputation is limited, trusting the action of third parties towards her (i.e. imitating) does favor a higher collective cooperation compared to random-trusting and share-alike mechanisms. We believe these results shed new light on the factors favoring social imitation as an adaptive mechanism in populations of cooperating social actors.

Introduction

The evolution of cooperative behavior in biological and human populations has been shown to rely critically on different forms of reciprocity [1–6]. In human society, cultural transmission mechanisms such as language allow for a subtle cooperative structure based on the principle of indirect reciprocity. In the absence of previous direct interactions which can be used to judge an individual, it is possible to observe and record the interactions of that individual with third parties [7,8], and assign a reputation to the individual guided by the principle: if I scratch your back, someone else will scratch mine [9]. Simulation models in which a reputation score associated with each actor records previous decisions about whether to cooperate or not, have revealed that indirect reciprocity among actors in a population will emerge particularly if all individuals have access to the reputation scores of other individuals [10–12].

However, when information about the past record of other individuals is unavailable or unreliable, laboratory experiments [13–20] and simulation models [14,21–23] have shown that actors might rely instead on imitation mechanisms or recognition heuristics to share resources with other actors they interact with according to their counterpart's trustworthiness. In fact, recent work has shown that cooperative behavior can spread as an imitation and trust mechanism across a population of self-regarding individuals [24]. The trustworthiness can be assigned to actors on the basis of how many third parties signal that they endorse a given actor, and as such is used as a proxy for the attributes of an individual when there is no detailed record of how those actors have acted towards others in the past [17,18,25,26]. This is to say, an actor C will extend trust to A (i.e. cooperate with A), because B previously extended trust to A, and in the absence of further information the trustworthiness of A can be used as part of a frugal heuristic or referral mechanisms by C [20,24,27,28].

Although reliance on imitation strategies can provide a heuristic that allows the identification of potentially trustworthy partners in interactions, there can be a negative impact on overall welfare since the resulting distribution of resources can reflect the principle of cumulative advantage [25,29–31]. Following this principle implies that the distribution of resources across actors in a population becomes increasingly skewed over time, with the rich getting richer and the poor getting poorer. Similarly, imitation strategies have proved extremely successful when applied to competitive strategies [32,33]. Here we explore how actors use different assessment attributes based on imitation and indirect reciprocity mechanisms to decide whom they cooperate with, and

who gains resources when repeated interactions are unlikely. We answer the questions of whether imitation and indirect reciprocity mechanisms can co-exist and generate collective cooperation, and whether imitation provides a reliable alternative to indirect reciprocity when information about an actor's reputation is frequently unavailable. In general, why is imitation a recurrent mechanism in human behavior given its potential negative effects on the distribution of resources in a population?

Results

The model

In our imitation-reciprocity (IR) model, we consider individuals faced with a social dilemma [1,9,34,35], who follow cooperative or altruistic strategies involving both reciprocity and imitation mechanisms [24,27,28]. Here, the donor has the opportunity to help a randomly chosen recipient at cost c, while the recipient gets a benefit b. Otherwise, the donor and recipient remain with their current payoff [10]. Hence, the donor faces a dilemma about whether to cooperate or not. However, we assume that non-cooperative actions harm the reputation and trust of the donor and recipient respectively. For reputation, we follow the image-scoring mechanism proposed in [10], where the image r of a donor is continually assessed according to their previous cooperative or non-cooperative actions towards other possible recipients in the population. Similarly, for imitative trust, a recipient's image t is continually assessed according to the cooperative or non-cooperative actions received from possible donors. Hence, the trust score of a recipient only records information about the action of third parties towards her. Donors have their own assessment strategies T_i and R_i for trust and reputation images respectively. A positive image of a recipient j always will make it more likely that a donor i will help than a negative image given by $t_j \geq T_i$ or $r_j \geq R_i$. This corresponds to the behavior of actors who have access to the reputation of potential recipients, and social actors using imitative strategies, who only have access to or are influenced by the trustworthiness of such recipients. The access to information is given by a threshold parameter p, which determines whether donors evaluate the reputation, with probability $1-p$, or the trustworthiness of recipients, with a probability p. In our

simulations, we consider n actors, which are replaced at the end of each generation m and transmit their strategies to the new population in proportion to their accumulated payoffs (Methods). In each generation, N randomly pair-wise interactions are chosen, where actors can play either the role of donors or recipients, i.e. $2N/n$ interactions per actor (see Methods for a detailed description of the IR model).

Imitation and indirect reciprocity

First, we analyze the effects of using imitative trust as an alternative mechanism to indirect reciprocity. We find that the collective payoff generated by indirect reciprocity is surprisingly robust to high levels of imitation. As illustrated in Figure 1A, we find that for most of the simulated levels of imitation $p \leq 0.7$, the average payoff per actor, calculated across the generation once the population has fixated into a common strategy, is higher than half of the maximum possible (i.e. payoff = 45). However, the average payoff considerably decreases as imitation becomes the only strategy followed by actors (i.e. payoff ≈ 1), revealing the elusive nature of cooperation. Similarly, analyzing the fixated strategies reached by the population in the last generation, we find that on average both imitative T (solid red line) and indirect reciprocity R (blue dashed line) strategies become non-cooperative (>0) at a high level of imitation $p = 0.8$ (see Fig. 1B). This suggests that only when imitation is used less than 80% as an assessment strategy, cooperative behavior dominates and the population achieve higher payoffs.

Additionally, we explore to what extent trusting the actions of others provides better cooperative outcomes than plausible alternative strategies. Our first alternative strategy or null hypothesis is a random-trusting process, where we assume that donors apply a simple probabilistic rule and cooperate on average 50% of the time. This is to say, when information about a recipient's reputation is unavailable p percent of the time, donors apply a simple random process and cooperate on average one out of two opportunities. For the second null hypothesis, we assume that donors follow a share-alike behavior, where they try to distribute benefits equally among all members in the population [36,37]. Here, donors cooperate if the trustworthiness (i.e.
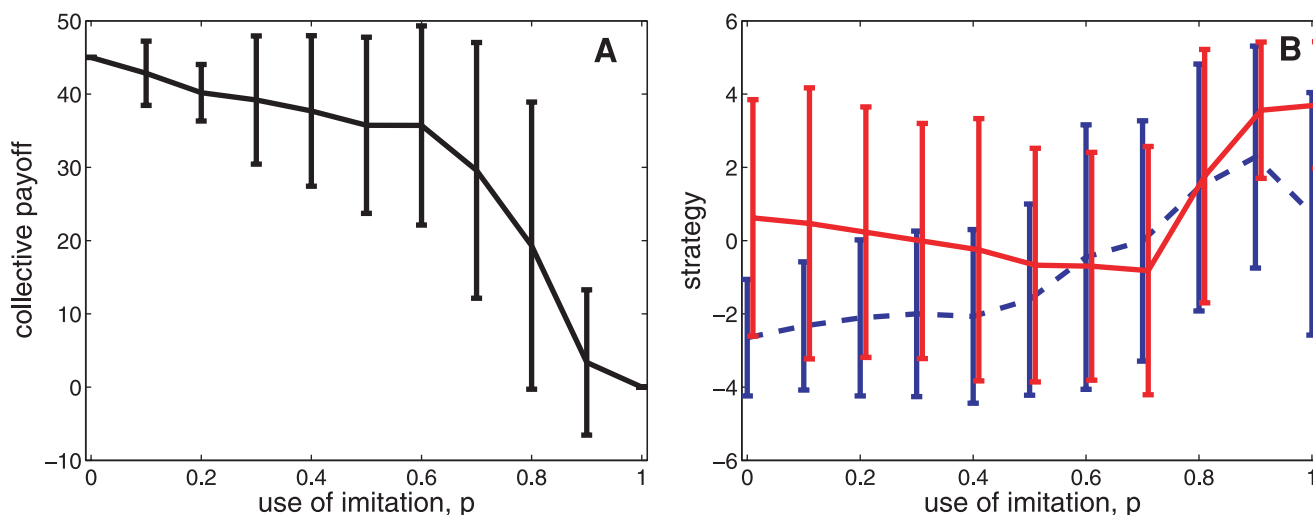

Figure 1. Imitative trust and indirect reciprocity. Panel **A** and **B** show, respectively, the average payoff per actor and the average strategies T (solid red line) and R (dashed blue line) observed in the population across different levels of imitation p (x-axis). Bars correspond to 2 standard deviations. Values are calculated over 1000 simulations considering the generation when the population has reached a fixated common strategy T, R.
doi:10.1371/journal.pone.0013475.g001

previous granted cooperation) of the recipient is low and defect if the trustworthiness is high (see Methods for details). Figure 2 shows that under intermediate levels of limited information $0.2 \leq p \leq 0.7$, random-trusting (green dashed line) processes and share-alike mechanisms (orange dashed line) display on average lower payoffs than imitative trust (black line). This reveals that trusting the action of others could be a useful alternative mechanism to indirect reciprocity when donors do not have frequent access to the reputation of potential recipients.

Vulnerability of cooperative strategies

To examine the vulnerability of different strategies on distributing equal number of resources to all members in the population, we investigate the effect of noise in the allocation of resources. We measure the distribution of payoffs in the population generated by changing the parameter p and introducing small errors in the decision-making process of donors [11]. This noise in the allocation of resources takes into account important effects such as memory constraints, bias in judgments or implementation errors [38–42]. We implement this by allowing donors to randomly change their decision with a small, fixed probability ϵ [11]. Here we consider that one out of ten times a donor can make an implementation or decision error ($\epsilon = 0.1$). Smaller values of ε generate similar results. Note that without this noise we would expect all actors with the same amount of resources.

To measure the distribution of payoffs in the population, we use the Gini coefficient [43]. The Gini coefficient represents the average difference in wealth share for two actors in the population normalized to fall between 0 (perfect equality) and 1 (maximum inequality). The Gini coefficient is defined as $G = \sum_{i=1}^{n} \sum_{j=1}^{n} |q_i - qj|/2n$, where $q_i = u_i / \sum_{k=1}^{n} u_k$, and u_i is the payoff of actor i, and n is the total number of actors in the population. Since payoffs can be negative, we take the minimum value as the baseline payoff equal to 1 and adjust all other payoffs accordingly. Figure 3A shows that under pure
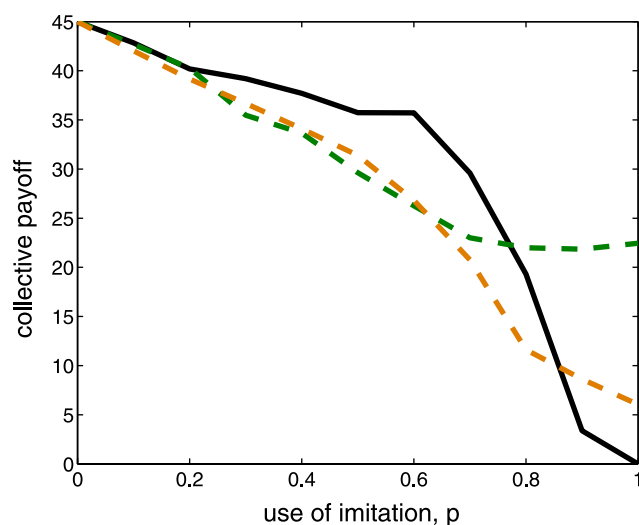
indirect reciprocity ($p = 0$), the population always favors cooperative strategies $R \leq 0$ with low Gini coefficients. Interestingly, Figure 3B shows that even when imitation and indirect reciprocity mechanisms are used equally $p = 0.5$, the population has a high likelihood ($\approx \%70$) of converging into cooperative strategies with low Gini coefficients (bottom left corner). By contrast, Figures 3C–D show that cooperators disappear and high Gini coefficients emerge at the point when imitation dominates the assessment mechanism in the population. Note that the highest Gini coefficients are reached when the population follows a trust-based cooperative strategy $T = 0$ combined with a reputation-based unconditional defector strategy $R = 6$. This shows that populations that only cooperate using imitation mechanisms are highly prone to inequality effects [25,29–31].

Emergence of cooperation

Finally, we investigate whether imitative trust and indirect reciprocity can co-exist and allow the emergence of reciprocal altruism. We explore how cooperation would evolve through mutations in a population of unconditional defectors. For each actor, we introduce a third dimension p. For simplicity we assume it can take three different values 0, 0.5 and 1, which correspond to the proportion of imitative trust used by actors, i.e. this dimension replaces the probability of using imitative trust caused by limited information in our original model (see Methods). We initialize the population with all actors having $p = 0$, $R = 6$ and a random strategy T, i.e. at the beginning actors only use indirect reciprocity strategies defined by unconditional defectors. To investigate the evolution of imitative trust and cooperation, we include mutations in the creation of new generations. We assume a small probability $h = 0.001$ [10] that a new actor adopts a randomly chosen strategy (T, R and p) than the one inherited by her parent.

First, our simulations show endless cycles of collective cooperation and defection. Figure 4A shows that the average payoff per actor per generation continuously fluctuates between 45, the maximum value, and 0, the minimum value. Second, we observe that imitative trust can, in fact, co-exist with indirect reciprocity. Figure 4B shows the percentage of actors with either $p = 0$ (blue), $p = 0.5$ (green) and $p = 1$ (red) across thousands of generations for a single simulation. The population continuously fluctuates between all the different strategies. Similar results hold if we only use $p = 0$ and $p = 1$. Note that the population never settles in a stable strategy. These results suggest that cooperation emerges only if indirect reciprocity is present; however, once this requirement is fulfilled, imitation can provide a plausible alternative strategy.

Discussion

It has been argued that cultural transmission mechanisms make it possible to assign a measure of reputation or social status to specific individuals in a population, so that cooperation can emerge in human societies as a consequence of indirect reciprocity [9,10]. However, the effects of reputation and social status do not necessarily coincide, and therefore are likely to warrant separate treatment. The evaluation of reputation takes into account the record of past actions of an individual, while social status reflects social preferences and mechanisms such as copying the helping behaviour of others [44]. Social experiments have shown that the actions and opinions of others used as a proxy for quality or reputation can affect someone's popularity or commitment to cooperate [18,24]. If we assume that actors are

Figure 2. Alternative cooperative mechanisms to imitation. The figure compares the average payoff per actor (see Fig. 1A) generated by imitative trust (black line) against the average payoff obtained by replacing imitative trust with random-trusting (green dashed line) and share-alike (red dashed line) mechanisms. Values are calculated over 1000 simulations considering the generation when the population has reached a fixated common strategy T, R.
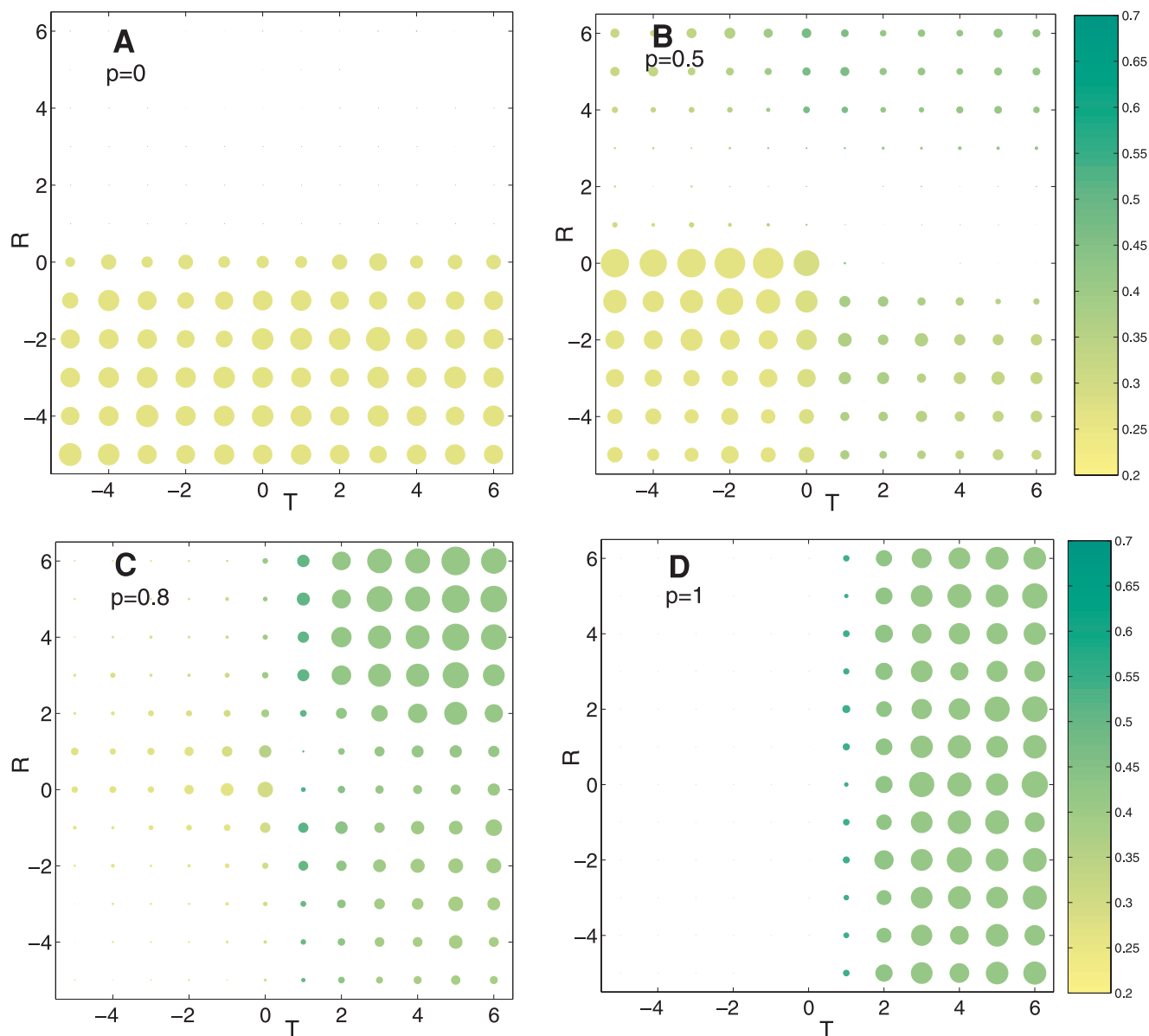doi:10.1371/journal.pone.0013475.g002

Figure 3. Vulnerability of cooperative strategies. We analyze the vulnerability of imitation and indirect reciprocity strategies on distributing similar resources when actors are subject to implementation errors. We introduce a probability $\varepsilon = 0.1$ that donors mistakenly act in the opposite way as it was expected from their strategy. Note that without errors we would expect all actors with the same amount of resources. Panels **A–D** show the correlation between Gini coefficients (shades) and the frequency of fixated strategies (circles) for $p = 0$, $p = 0.5$, $p = 0.8$ and $p = 1$ respectively. Gini coefficients and frequencies are reported as the average over 10^5 simulations considering the generation when the population has reached a fixated common strategy T, R. The frequency of occurrence for each strategy is proportional to the area of the circles.
doi:10.1371/journal.pone.0013475.g003

heterogeneous in this regard, then it is useful to model populations so that individuals vary in how they attend to these two types of information, and hence to allow for different combinations of reputation and imitative trust mechanisms. Although there is no difference in principle with regard to the cognitive demands imposed by each mechanism, there may also be asymmetries between the availability or quality of information associated with giving help and receiving help in a given social setting. Methodologically, the addition of imitative trust to the original indirect reciprocity model restores balance to how information on donors and recipients is treated. Each pairwise interaction between a donor and recipient encodes information

about both parties, which the combination of image and imitative-trust scores fully captures.

Here we have analyzed for the first time the effects that two assessment mechanisms –imitation and indirect reciprocity, which determine the structure of who cooperates with whom and who gains resources, might generate when access to the reputation of potential recipients is frequently unavailable or actors are influenced by the cooperative action of others. We have found that both the cooperative behavior and the fair allocation of resources decrease as the use of imitation mechanisms increases. However, we have also found that as long as actors use imitation and indirect reciprocity mechanisms equally, cooperative strategies
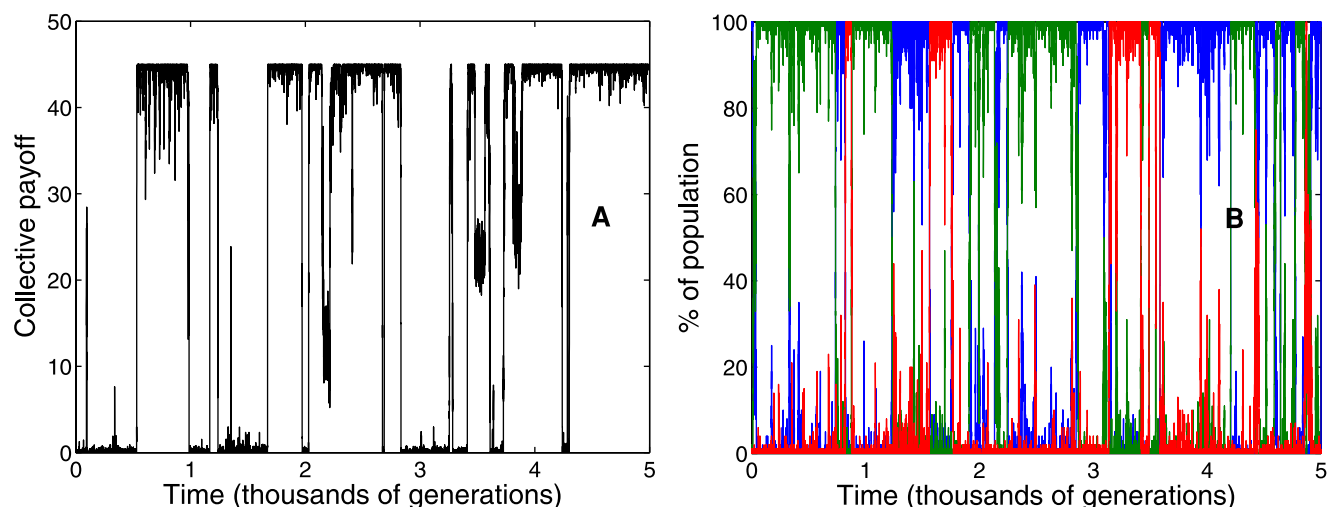
Figure 4. Emergence of cooperation. To investigate the emergence of cooperation and the co-existence of imitation and indirect reciprocity, we consider that new actors will adopt a randomly chosen strategy with probability $h = 0.001$ (see text). Additionally, to differentiate between actors using only indirect reciprocity, imitation or a mix of the two, we introduce a third dimension p that takes values of 0 (blue line), 1 (red line), or 0.5 (green line) respectively. We initialize the population with $p = 0$ and $R = 6$, i.e. unconditional defectors. Panel **A** shows the average payoff per actors per generation for a single simulation. Note that the population continuously fluctuate between maximum cooperation and defection. Panel **B** shows that the strategies p also fluctuate across generations. This reveals that although there is no stable strategy, actors can adopt cooperative imitative and indirect reciprocity strategies.
doi:10.1371/journal.pone.0013475.g004

dominate and resource inequalities are small. Surprisingly, we have observed that trusting the action of others generates higher payoffs than simple random-trusting processes and share-alike mechanisms. This suggests that imitation might be in fact an adaptive mechanism in populations of cooperating social actors under limited information.

Materials and Methods

IR Model

We study reciprocal altruism under imitative trust and indirect reciprocity. Specifically, we consider n actors over a fixed lifetime, which are replaced at the end of each generation m. In each generation, N randomly pair-wise interactions are chosen, where actors can play either the role of donors or recipients (i.e. $2N/n$ interactions per actor). If a donor i cooperates with a recipient j, the donor pays a cost c and the recipient gets a benefit b. Otherwise, if the donor does not cooperate, both payoffs remain exactly the same. A donor i decides whether to cooperate or not based on the recipient's image and her own assessment strategy. The image of a recipient j is assessed either by her trust score t_j or reputation score r_i, where both can take integer values in $[-5, 5]$ following the standard convention of reference [10]. A tunable parameter p gives the probability that the donor evaluates the recipient's trust score and with probability $1 - p$ the donor evaluates the recipient's reputation score. In addition, a donor i has her own assessment strategies T_i and R_i, drawn from a uniform distribution in $[-5, 6]$, for trust and reputation respectively. Therefore, the model comprises 144 different strategies. According to whether the donor evaluates the recipient's trustworthiness or reputation, cooperation will be established if the recipient's image is above a certain threshold given by $t_j \geq T_i$ or $r_j \geq R_i$ for trustworthiness and reputation respectively. If cooperation is established, the donor's reputation r_i is increased by one unit, else her reputation decreases by one unit. In addition,

each time the recipient receives cooperation her trustworthiness t_j is increased by one unit, else her trustworthiness decreases by one unit. Note that the increase and decrease of scores is subject to the boundary conditions of the score values $[-5, 5]$. This score boundary allows the presence of unconditional cooperators $T = R = -5$ and unconditional defectors $T = R = 6$. At the end of its lifetime, the population is replaced by a new generation, where an old actor i can transmit her assessment strategies T-R to a new actor k, with a probability w_i proportional to her own payoff and relative to the payoffs of all actors j in the population [45]. Mathematically, this is given by $w_i = u_i / \sum_j^n u_j$, where u_i is the payoff of actor i. Since payoffs can be negative, we take the minimum value as the baseline payoff equal to 1 and adjust all other payoffs accordingly. If not stated otherwise, all generations start with $t_i = r_i = u_i = 0$ for all actors i. Simulations were performed using conventional parameter values [10,11]: $m = 500$, $N = 500$, $n = 100$, $c = 1$ and $b = 10$. We also extended our model for large populations with up to $n = 10^6$ actors and found similar results.

Share-alike mechanism

According to whether the donor evaluates the recipient's trust or reputation scores, cooperation will be established if the recipient's image is below a certain threshold given by $t_j \leq T_i$ or $r_j \geq R_i$ for trust and reputation respectively.

Acknowledgments

Author Contributions

Conceived and designed the experiments: SS DS FRT. Performed the experiments: SS DS. Analyzed the data: SS DS FRT. Wrote the paper: SS DS FRT.

References

1. Ostrom E, Burger J, Field CB, Norgaard RB, Policansky D (1999) Revisiting the commons: Local lessons, global challenges. Science 248: 278–282.
2. Axtell RL (2001) Zipf distribution of u.s. firm sizes. Science 293: 1818–1820.
3. Fehr E, Fischbacher U (2003) The nature of human altruism. Nature 425: 785–791.
4. Hammerstein PE (2003) Genetic and Cultural Evolution of Cooperation. MIT Press.
5. Roberts G (2007) Evolution of direct and indirect reciprocity. Proceedings of the Royal Society B 275: 173–179.
6. Saavedra S, Reed-Tsochas F, Uzzi B (2009) A simple model of bipartite cooperation for ecological and organizational networks. Nature 457: 463–466.
7. Wedekind C, Milinski M (2000) Cooperation through image scoring in humans. Science 288: 850–852.
8. Basu S, Dickhaut J, Hecht G, Towry K, Waymire G (2009) Recordkeeping alters economic history by promoting reciprocity. Proc Natl Acad Sci of USA 106: 1009–1014.
9. Alexander RD (1987) The Biology of Moral Systems. Aldine de Gruyter.
10. Nowak MA, Sigmund K (1998) Evolution of indirect reciprocity by image scoring. Nature 393: 573–577.
11. Leimar O, Hammerstein P (2001) Evolution of cooperation through indirect reciprocity. Proc R Soc B 268: 745–753.
12. Pacheco JM, Santos FC, Chalub FACC (2006) Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. PLoS Computational Biology 12: e178.
13. Pingle M (1995) Imitation versus rationality: An experimental perspective on decision making. J of Socio-Economics 24: 281–315.
14. Bikhchandani S, Hirshleifer D, Welch I (1998) Learning from the behavior of others: Conformity, fads, and informational cascade. The J of Economic Perspectives 12: 151–170.
15. Gergely G, Bekkering H, Király I (2002) Rational imitation in preverbal infants. Nature 415: 755.
16. Goldstein DG, Gigerenzer G (2002) Models of ecological rationality: The recognition heuristic. Psychological Rev 109: 75–90.
17. Cialdini RB, Goldstein NJ (2004) Social influence: compliance and conformity. Annual Rev Psych 55: 591–621.
18. Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. Science 311: 854–856.
19. Tomasello M (2008) Origins of Human Communications. MIT Press.
20. Boero R, Bravo G, Castellani M, Squazzoni F (2009) Reputational queues in repeated trust games. The Journal of Socio-Economics 38: 871–877.
21. Banerjee AV (2002) A simple model of herd behavior. The Quarterly J of Economics 107: 797–817.
22. Bolton GE, Katok E, Ockenfels A (2005) Cooperation among strangers with limited information about reputation. J of Public Economics 89: 1457–1468.
23. Na JP, Volken H, Pestelacci E, Tomassini M (2009) Conformity hinders the evolution of cooperation on scale-free networks. Phys Rev E 80: 016110.
24. Fowler JH, Christakis NA (2010) Cooperative behaviour cascades in human social networks. Proc Natl Acad Sci of USA 107: 5334–5338.
25. Merton RK (1968) The matthew effect in science. Science 159: 56–63.
26. Podolny JM (2005) Sociological Signals: A Sociological Study of Market Competition Princeton Univ. Press.
27. Ostrom E, Walker J (2003) Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research Russell Sage Foundation Publications.
28. Seinen I, Schram A (2006) Social status and group norms: Indirect reciprocity in a repeated helping experiment. European Economic Review 50: 581–602.
29. Zipf GK (1942) The unity of nature, least-action and natural social science. Sociometry 5: 48–62.
30. de Solla Price DJ (1976) A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science 27: 292–306.
31. Ijiri Y, Simon HA (1977) Skew Distributions and the Sizes of Business Firms North-Holland.
32. Axelrod R, Hamilton WD (1981) The evolution of cooperation. Science 211: 1390–1396.
33. Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K, et al. (2010) Why copy others? insights from the social learning strategies tournament. Science 328: 208–231.
34. Bowles S (2008) Policies designed for self-interested citizens may undermine the moral sentiments: Evidence from economic experiments. Science 320: 1605–1609.
35. Gintis H, Bowles S, Boyd RT, Fehr EE (2005) Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life. MIT Press.
36. Fehr E, Bernhard H, Rockenbach B (2008) Egalitarianism in young children. Nature 454: 1079–1084.
37. Tomasello M, Warneken F (2008) Share and share alike. Nature 454: 1057–1058.
38. Milinski M, Wedekind C (1998) Working memory constrains human cooperation in the prisoners dilemma. Proc Natl Acad Sci USA 995: 13755–13758.
39. Mohtsuki H (2004) Reactive strategies in indirect reciprocity. J Theor Biol 227: 299–314.
40. Brandt H, Sigmund K (2006) The good, the bad and the discriminator errors in direct and indirect reciprocity. J Theor Biol 239: 183–194.
41. Sommerfeld RD, Krambeck HJ, Semmann D, Milinski M (2007) Gossip as an alternative for direct observation in games of indirect reciprocity. Proc Natl Acad Sci of USA 104: 17435–17440.
42. Rankin DJ, Eggimann F (2009) The evolution of judgement bias in indirect reciprocity. Proc R Soc B 276: 1339–1345.
43. Gini C (1921) Measurement of inequality of incomes. The Economic J 31: 124–126.
44. Lehmann L, Foster KR, Borenstein E, Feldman MW (2008) Social and individuals learning of helping in humans and other species. Trends in Ecology and Evolution 23: 664–671.
45. Hofbauer J, Sigmund K (1987) Evolutionary Games and Population Dynamics Cambridge Univ. Press.