THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

LSE Research **Online**

Heinrich H. Nax, Matjaž Perc, Attila Szolnoki & Dirk Helbing

Stability of cooperation under image scoring in group interactions

Article (Published version)
(Refereed)

This version available at: http://eprints.lse.ac.uk/65431/

Available in LSE Research Online: Online: February 2016

http://eprints.lse.ac.uk

SCIENTIFIC REP☿RTS

Stability of cooperation under image scoring in group interactions

Heinrich H. Nax[1], Matjaž Perc[2,3], Attila Szolnoki[4] & Dirk Helbing[5,6]

Image scoring sustains cooperation in the repeated two-player prisoner's dilemma through indirect reciprocity, even though defection is the uniquely dominant selfish behaviour in the one-shot game. Many real-world dilemma situations, however, firstly, take place in groups and, secondly, lack the necessary transparency to inform subjects reliably of others' individual past actions. Instead, there is revelation of information regarding groups, which allows for 'group scoring' but not for image scoring. Here, we study how sensitive the positive results related to image scoring are to information based on group scoring. We combine analytic results and computer simulations to specify the conditions for the emergence of cooperation. We show that under pure group scoring, that is, under the complete absence of image-scoring information, cooperation is unsustainable. Away from this extreme case, however, the necessary degree of image scoring relative to group scoring depends on the population size and is generally very small. We thus conclude that the positive results based on image scoring apply to a much broader range of informational settings that are relevant in the real world than previously assumed.

Public goods provision[1,2], common-pool resource management[3], and other social dilemma situations often require large groups of individuals to make individually costly contributions toward a collective action, i.e., to cooperate[4]. Without sufficient consideration of future consequences, however, many real-world interactions lack the necessary cooperativeness to prevent misalignment of private and social interests. This leads to socially undesirable outcomes. The result may be irrevocable mismanagement and over-exhaustion of shared resources, ultimately resulting in the 'tragedy of the commons'[3,5,6], which is a long-term outcome that is worse for everyone.

Hence, the 'puzzle of cooperation'[7–11]. Why do certain social interactions flourish with high levels of foresight and cooperation, while others are impeded by short-sighted free-riding behaviour? Game theory[12,13] provides the necessary theoretical framework for studying the individual-level motivations[14] to answer this question. Indeed, game theory provides many answers, including a variety of evolutionary[15,16] and psychological[17–19] explanations, but only two based, in the language of game theory, on 'rationality'.

One of these two rational explanations is based on foresightedness. Namely, if individuals take future consequences sufficiently into account, then equilibria exist supported by more sophisticated, repeated-game strategies that overcome the short-term incentives to free-ride[20]. Unfortunately, to uphold cooperation that way, individuals would need to commit to strategies in ways that may produce grim consequences if not everyone follows the strategy equilibrium path[21]. Hence, groups consisting of sufficiently foresighted agents may succeed to guarantee provision of a common, while groups with too many short-sighted or boundedly rational agents will fail. The second rational explanation of the puzzle of cooperation is based on building a reputation from the past[22]. Essentially, players build a commitment-to-cooperation reputation through their past actions, and find each other, thus, outperforming defectors (who are stuck with other defectors), even though defection is the uniquely dominant

[1]Department of Social Sciences, ETH Zürich, 8092 Zürich, Switzerland. [2]Department of Physics, Faculty of Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. [3]Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, SI-2000 Maribor, Slovenia. [4]Institute of Technical Physics and Materials Science, Centre for Energy Research, Hungarian Academy of Sciences, P.O. Box 49, H-1525 Budapest, Hungary. [5]Computational Social Science, ETH Zürich, 8092 Zürich, Switzerland. [6]Risk Center, ETH Zürich, 8092 Zürich, Switzerland. Correspondence and requests for materials should be addressed to H.H.N. (email: hnax@ethz.ch)

selfish behaviour in the one-shot game[14]. As a result, if reputation matters sufficiently in determining with whom individuals interact, then cooperation can survive despite limited foresight.

Unfortunately, due to the inherent simultaneity of interactions[23], it is impossible to condition one's own decision on the decisions of the others[14]. One of the most important, and perhaps the simplest, reputation mechanism known in the literature to overcome this problem is image scoring[24]. Famously, image scoring can sustain cooperation in the repeated two-player prisoner's dilemma through various forms of indirect reciprocity[25–27]. Under image scoring, agents learn who cooperated and who defected in previous interactions, and consequently condition their own actions on this information. Essentially, image scoring enables cooperators to find each other, and this overcomes the negative Nash equilibrium prediction of universal defection from the one-shot game. Interestingly, image scoring has been shown to work in the laboratory[28–32], but in general, it is considered to provide a relatively frail support to cooperative behaviour[33,34].

In fact, many real-world social dilemmas unfold in groups[35], and it is unlikely that individuals will have access to others' individual action histories[1,2]. Information that should be readily available, however, concerns the performance of the groups as a whole. Such information thus enables 'group scoring' as an alternative to image scoring. In particular, the image of an individual is no longer determined by its own past action, but by the performance of the group where an individual is member. More precisely, each player's group score summarizes the aggregate cooperativeness of the groups where he was a member in the past, without any additional information regarding what the player did individually. Two important and previously unaddressed new questions emerge: (i) How do results related to image scoring generalize to group scoring?, and (ii) How sensitive are these results to information, that is, when image scoring constitutes a proportion $p \in [0,1]$ of information made available and the residual information is based on group scoring? The common feature of previous research on image scoring is that, over time, cooperators achieve higher scores and defectors achieve lower scores, and interactions are matched based on these scores such that thus cooperators play with cooperators and defectors play with defectors. In our paper, we build on this common feature by assuming the existence of a mechanism that assorts and matches players by their scores. In addition, we extent the scope of image-scoring-based models by analysing the sensitivity of the results to the imperfections of scores that ought to reflect individuals' true past cooperativeness rather than the overall performance of the groups where they are members. Our analysis continuously spans the worlds of two extremes; image scoring and group scoring. As we will show, image scoring, reflecting accurately individuals' past actions, works perfectly also in the generalized prisoner's dilemma game that is governed by group interaction. Conversely, group scoring fails, as it enables defectors to effectively hide behind the cooperative efforts of others in the group. But how many true images are needed for cooperation to evolve in group interaction? In other words, what is the necessary proportion $p \in (0,1)$ of image scoring? It turns out that this depends sensitively on the underlying parameters of the interactions in ways that provide a formal basis for some of Ostrom's conditions for successful common-pool resource management[3]. Key determinants are the rate of return, the size of the population, and the group size. Remarkably, for large populations only a 'grain' of image scoring is generally sufficient for cooperation to become dominant.

Results

We shall now formalize our arguments, generalizing step by step the two-by-two prisoner's dilemma model due to[24] to the more general context of the voluntary contributions game[1,2]. Suppose a population $N = \{1, 2, \ldots, n\}$ plays the following game in rounds $t = \{1, 2, \ldots, \infty\}$. Each player $i \in N$, in period t, chooses a contribution c_i^t from budget $B = \{0, 1\}$. At the same time, each player i is associated with a binary *score* $s_i^t = \{0, 1\}$, and players are matched into k groups $\mathcal{S} = \{S_1, \ldots, S_k\}$ of a fixed size $s = n/k$ according to the ranking of players' scores (with random tie-breaking). After groups have formed, i's resulting payoff turns out $\phi_i^t(c) = (1 - c_i^t) + r/s * \sum_{j \in S_i^t} c_j^t$, where S_i^t is i's group. r is the game's fixed rate of return, and r/s the game's 'marginal per-capita rate of return' that summarizes the underlying game's synergy. We assume, as is standard, $r \in [1, s]$, that is, contributing a unit of budget is socially beneficial (yielding a sum total of payoffs to all players larger than one), but individually costly.

We consider the following range of scoring mechanisms between image scoring and group scoring.

Image scoring. First, we formulate the equivalent of *image scoring*[24] in our setup: at time t each player i has an image score, s_{Ii}^t, known to every player which is based on decisions prior to t. In period $t + 1$, if i's period-t contribution c_i^t exceeded the average contribution $\bar{c}^t = \sum_{i \in N} c_i^t / n$ at time t, then his image score is one; if $c_i^t < \bar{c}^t$ then it is zero. Finally, if $c_i^t = \bar{c}^t$, then $s_i^{t+1} = s_i^t$.

Group scoring. Analogously, we formulate *group scoring*: at time t each player i has a group score, s_{Gi}^t. Suppose i's period-t group is S. If the contribution in S, $c_S^t = \sum_{i \in S} c_i^t$, exceeded the average group contribution overall, $\bar{c}_S^t = \sum_{S \in \mathcal{S}^t} c_S^t / k$, then i's period-$(t+1)$ group score is one; if $c_S^t < \bar{c}_S^t$ then it is zero. Finally, if $c_S^t = \bar{c}_S^t$, then $s_i^{t+1} = s_i^t$.
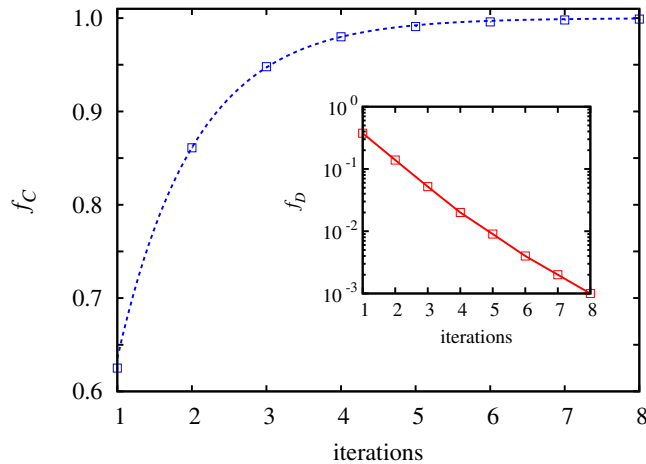
Figure 1. Under image scoring cooperators always rise to complete dominance. Shown is the fraction of cooperators in dependence on the number of iterations. The inset shows the corresponding fraction of defectors on a semi-logarithmic scale to highlight their exponential decay. Depicted results are averaged over 200 independent realizations to minimize fluctuations. Parameter values are: $s = 5$, $n = 1500$ and $r = 4$.

Hybrid scoring. A hybrid between *image scoring* and *group scoring* in our setup means that, at time t, each player i has a hybrid score, s_{Hi}^t, known to every player which is based on decisions prior to t. In period $t + 1$, if i's score is updated according to image scoring with probability p, and according to group scoring with probability $1 - p$.

To summarize the three scoring methods, the types of information necessary under the different regimes are as follows. For image scoring, ex post individual-level information about contribution decisions is necessary, group-level information is therefore trivially also available. For group scoring, group associations and ex post group-level contributions need be known, individual-level information is not necessary. For the hybrid case, characterized by degree of image scoring $p \in [0, 1]$, the probability that individual-level information rather than only group-level information becomes available must be larger than zero.

The stability of cooperation under the different scoring rules can be evaluated. One result is that tragedy of the commons (resulting from universal defection) is a potential risk in all cases. The stability of universal defection under all scoring rules derives from the fact that unilateral defection is a best response under all scoring methods against a state of universal defection. However, the relative stability of this worst-case outcome vis-a-vis a highly cooperative state critically depends on the scoring rule, mitigating this issue.

Under image scoring, high levels of cooperation can be stabilized and then turn out to be more stable. This is the case if the proportion of cooperators with score one (matched in good groups) grows exactly at the speed so as to neutralize the shrinking of the proportion of cooperators with score zero (matched in bad groups). The defectors profit from the contributions of the latter group and achieve an average growth equal to that of the average cooperator.

Stability is summarized with the results presented in Fig. 1 from simulations. The Methods section contains analytical proof of these results, as well as a description of the employed Monte Carlo simulation procedure. It can be observed that the state of complete cooperation is reached exponentially fast. We emphasize that this result is recovered independently of the value of s, r and n, and it is also robust against variations of the strategy adoption rule. Cooperation will always prevail under image scoring, as it allows cooperators to separate from defectors. In general, cooperators form homogeneous groups that provide them with a competitive payoff. Conversely, defectors must be content to form groups with their like, which provides them a null payoff. Cooperators can therefore easily invade defectors, and they do so with a speed that is proportional to their number, which ultimately gives rise to the exponentially fast downfall of defectors.

At first sight, such a state of cooperation may also seem a candidate for stability under group scoring. Inspection of the individual growth dynamics, however, reveals one crucial difference. Namely, cooperation states are not robust against the influx of defectors with score one. These players outperform all others, which, jointly with the fact that score-zero defectors outperform score-zero cooperators, implies an above-average growth rate for defection vis-a-vis cooperation. In other words, the key difference between image scoring and group scoring is that defectors can only free-ride on the contributions of others under image scoring, while, under group scoring, defectors can free-ride on the contributions *and* scores of others.
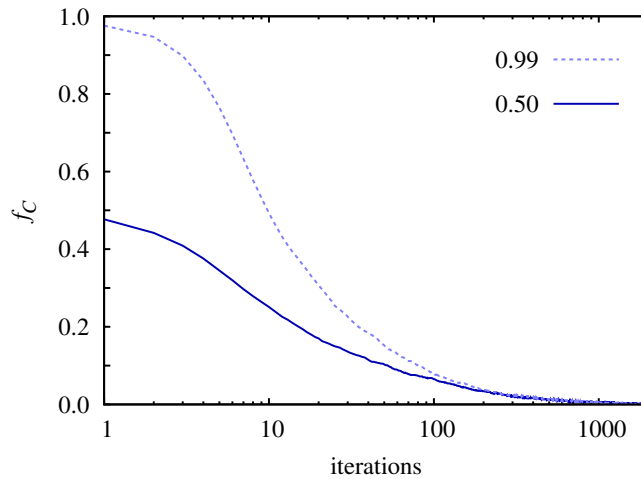
Figure 2. Under group scoring cooperators always die out. Shown is the fraction of cooperators in dependence on the number of iterations for two different initial fractions (see legend). Depicted results are averaged over 200 independent realizations to minimize fluctuations. Parameter values are the same as in Fig. 1.

Figure 2 mirrors the set-up of Fig. 1 in terms of the simulation procedure, only that here group scoring instead of image scoring is used. It can be observed that, irrespective of the initial fraction of cooperators, they eventually die out. As by image scoring, this outcome too is robust against variations of s, r, n and the strategy adoption rule. Group scoring allows defectors to have the same high score as cooperators, which in turn disables the separation of the two strategies into homogeneous groups. In agreement with the outcome of the public goods game in a well-mixed population, even a single defector can therefore eventually invade the entire population. Groups scoring thus completely fails to mitigate the tragedy of the commons.

Since image scoring and group scoring could not be more different in their ability to stabilize cooperation, it remains of interest to determine the merit of hybrid scoring. While it seems reasonable to assume that sometimes the information about the past of each particular individual is readily available, more often that not the scoring of an individual is possible only indirectly through the achievements of the groups where s/he was member. We note that individual contributions in group efforts are notoriously difficult to pinpoint, which is also why the reciprocation to such efforts is quite a vague concept — if a group contains a cooperator and a defector, who do you reciprocate with[36]? The question thus is, just how much individual-level information is needed to stabilize cooperation? To answer this question, we introduce the probability p that a player's score is determined by image scoring, while otherwise, with probability $1 - p$, group scoring is used. All other simulation details remain the same as in Figs 1 and 2.

Results presented in Fig. 3 show that cooperation can evolve even at a very small p value, if only the population size is sufficiently large. The key for the stability of cooperation is for cooperators being able to recognize each other through their high scores, and thus to form homogeneous groups. The lower the value of p and the lower the value of r, the longer it takes for cooperators to segregate from defectors. Since cooperators are threatened by extinction, it is imperative that the segregation occurs before defectors take over. Accordingly, the lower the value of p and r, the larger the population size needs to be to warrant sufficient time to cooperators to segregate before they die out. A lower bound for p is $p \geq 2/n$. Results presented in Fig. 4 make these arguments quantitatively more accurate. Evidently, the lower the population size, the higher the values of p and r need to be for cooperation to prevail. In small populations, there exist critical threshold values for both r (main panel) and r (inset), where drops to defector dominance are abrupt and occurring without precursors.

We emphasize that the results concerning hybrid scoring mechanisms are independent of the group size as long as their number, and hence the population size, is sufficiently large, and they are also independent of the strategy adoption rule. This corroborates our main argument, which is that, regardless of the scoring that is used, conditions need to be given for cooperators to completely segregate from defectors, i.e., to form homogeneous groups without a single defector. The identification of defectors has second-order importance. The key goal of scoring is thus to allow cooperators to recognize each other efficiently and to form homogenous groups accordingly.

Discussion

"The most important unanswered question in evolutionary biology, and more generally in the social sciences, is how cooperative behaviour evolved and can be maintained in human or other animal groups and societies" (Robert May in his Presidential Address to the Royal Society in 2005). A seminal
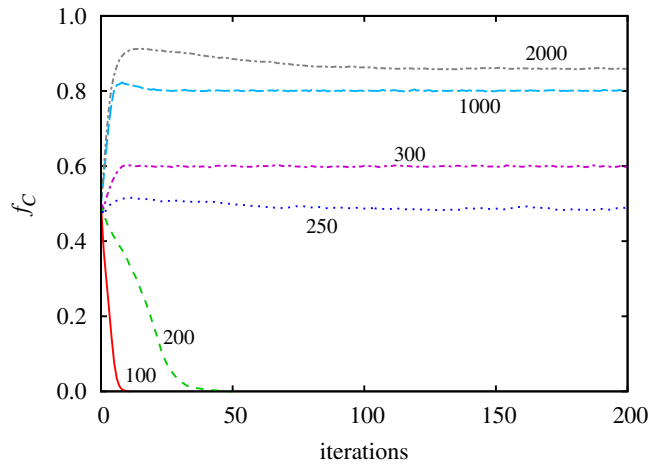
Figure 3. Under hybrid scoring, even a minute probability for accessing the image score of an individual player suffices to stabilize cooperation. However, the size of the population is crucial, as the segregation of cooperators from defectors takes the longer the smaller the value of *p*. Thus, if the population size is too small, cooperators are likely to die out before segregating into homogeneous groups. Segregation is also decelerated by low values *r* (see Fig. 4 for details). Shown is the fraction of cooperators in dependence on the number of iterations, as obtained for different numbers of groups forming the population (indicated alongside the lines). Depicted results are averaged over 50–200 independent realizations to minimize fluctuations. Parameter values are: $p = 0.01$, $s = 5$ and $r = 1.1$.

explanation for the puzzle of cooperation is based on image scoring[24], a mechanism that is both stunningly successful and stunningly simple. However, in its original formulation and application[24], it came with the restriction to interactions that are pairwise to informational environments that allow a complete tracking of individual-level information.

In the real world, these restrictive assumptions may not be germane. Instead, cooperation may involve several groups of individuals, and ex post information regarding individual-level cooperativeness may only percolate imperfectly through group-level information. Our focus in this paper has been to determine robust theoretical predictions regarding the emergence and survival of cooperation in such situations. The presented results have rather important implications. One is negative. Namely, when individual-level information is not available, cooperation cannot spread. But there is a silver lining. When there is at least a 'grain' of individual-level information, this may suffice for cooperators to find each other and form groups that are impervious to an invasion by defectors. We have shown that the spread of cooperation is robust to extensive imperfections in image scoring, thus extending the domain of environments where we should expect flourishing cooperation levels based on this established mechanism that fosters indirect reciprocity. Factors that affect the effectiveness of hybrid scoring negatively and positively are group size and population size, respectively. The rate of return was shown to not matter.

There is one important component that the model we considered here externalises, namely the issue of how a hierarchy of scores translates into an analogous group formation hierarchy. These are questions future work should address. In particular, we need to address how mechanisms known to play such a role in the context of image scoring would translate into the informational setting considered here, and how such mechanisms may be designed. Little is known in this direction. Economic experiments[37,38] might be particularly conductive to such research and guide future theoretical work to relevantly address these fundamental dilemmas of human cooperation.

Methods

Simulation procedure. The employed Monte Carlo simulation procedure[39] requires the iteration of the following three elementary steps. First, two randomly selected players *i* and *j* play one instance of the public goods game in their current group, thereby obtaining payoffs ϕ_i and ϕ_j, respectively. Next, player *j* adopts the strategy of player *i* with the probability given by the Fermi function $W = 1/\{1 + \exp[(\phi_j - \phi_i)/K]\}$, where $K = 0.1$ quantifies the uncertainty by strategy adoptions[40]. Each full Monte Carlo step gives a chance for every player to change its strategy once on average. The reported fractions of cooperators and defectors were determined in the stationary state.

Stability analysis. At a given time *t*, there are four action-score pairs: cooperate-one ($c_i^t = s_i^t = 1$), defect-one ($c_i^t = 0$ and $s_i^t = 1$), cooperate-zero ($c_i^t = 1$ and $s_i^t = 0$), and defect-zero ($c_i^t = s_i^t = 0$). Suppose players are hardwired to play either action. However, scores change. Depending on the scoring mechanism, the action vector *c* in period-*t* implies a score (or a probability distribution of scores) for
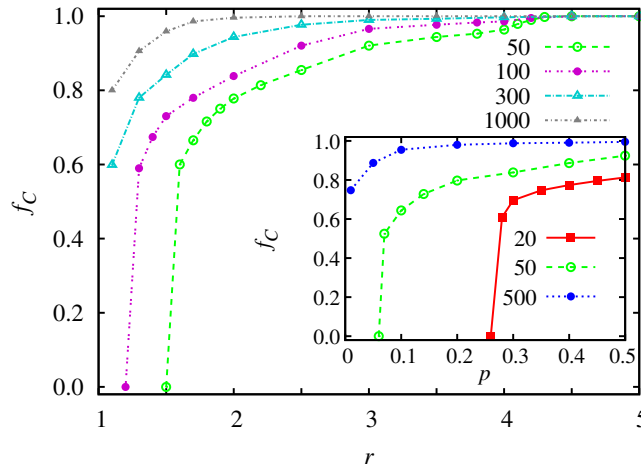
Figure 4. Hybrid scoring requires a critical population size to stabilize cooperation under adverse conditions. Cooperators essentially compete to segregate from defectors before being completely wiped out. The lower the value of r (main panel) and p (inset), the larger the population size needed for cooperation to be stabilized. Note that for a sufficiently small population size (see legend), there exist sudden drops to zero cooperation levels at critical values of r and p. Shown is the stationary fraction of cooperators in dependence on r (main) and p (inset), as obtained for different numbers of groups forming the population (see legend). Parameter values are: $p = 0.01$ (main) and $r = 1.1$ (inset).

period $t + 1$. Assume that in some period all four action-score pairs are represented by positive population proportions, $p_{C1}, p_{D1}, p_{C0}, p_{D0}$. Write p_C, p_D for the proportions of cooperators and defectors. Suppose the action proportions grow/ shrink, for action $a = \{C, D\}$, given by a replicator equation similar to that in the standard form: $\partial p_a / \partial t = p_{a1}^t * (\pi_{a1} - \overline{\pi}) + p_{a0}^t * (\pi_{a0} - \overline{\pi})$, where π_{as} is the expected payoff to action-score pair as and $\overline{\pi}$ is the average population payoff.

Stability of unconditional defection under all scoring rules. Suppose all players j in period t, independent of their score, defect except for one player i who plays $c_i^t = 1$. W.l.o.g., suppose $s_i^t = 1$. No matter what i's score, the payoff to i representing p_{C1} is $\pi_{C1} = r/s < 1$, while the average payoff is $\overline{\pi} = 1 + \frac{s-1}{n-1}(r/s) > 0$. Hence, $\partial p_C / \partial t < 0$ and therefore any such process is stable at $p_C = 0$.

Stability of high cooperation levels under image scoring. Suppose the four different strategies at time t have mass of $p_{C1}, p_{D1}, p_{C0}, p_{D0}$ respectively such that $p_{D1} = 0$. We shall now show that there exists a starting state with $p_C > (s-1)/s$ such that $\partial p_C / \partial t = 0$. Suppose that $p_{C1} = (n-s)/n$. Then $\pi_{C1} = r$, $\pi_{C0} = (s - n*p_{D0})*(r/s)$ and $\pi_{D0} = 1 + (s - n*p_{D0})*(r/s)$. $\partial p_C / \partial t = 0$ if $\frac{n-s}{n}*r + \frac{s - n*p_{D0}}{n}*(s - n*p_{D0})*(r/s) = p_{D0}*\left(1 + (s - n*p_{D0})*(r/s)\right)$. This yields $p_{DO} = 1/(1 - r + n*(r/s))$. Hence, a high cooperation level of $p_C = \frac{n*(r/s) - r}{1 - r + n*(r/s)}$ is stable if $p_{C1} = (n-s)/n$, $p_{D1} = 0$, $p_{C0} = (s/n) - 1/(1 - r + n*(r/s))$, and $p_{DO} = 1/(1 - r + n*(r/s))$. Notice that the temporary growth in p_{C1} relative to p_{C0} is compensated by the score dynamics which imply that exactly the proportion by which p_{C0} will be replaced by the growth in p_{C1}. Notice also that these proportions are robust to a small increase in $p_{D1} = 0$ as the overproportional growth of $D1$ will result in scores of zero in the next period (all $D1$s will turn $D0$ in the next round), and then lead to a shrinking of $D0$.

Instability of high cooperation levels under group scoring. At first sight, the state with $p_{C1} = (n-s)/n$, $p_{D1} = 0$, $p_{C0} = (s/n) - 1/(1 - r + n*(r/s))$, and $p_{DO} = 1/(1 - r + n*(r/s))$ could be stable. Indeed, provided that $p_{D1} = 0$ this state is stable against growth of proportions p_{C1}, p_{C0} and p_{DO}. However, any small increase in p_{D1} destabilizes the whole system. This is because, given p_{C1}, p_{C0} and p_{DO}, the non-zero proportion p_{D1} not only grows overproportional, but also — and this is the key difference with image scoring-- is likely to keep a score of one, and will therefore also not shrink in the next period. Hence, the growth of $C1$ slows down while $D1$ keeps growing. There can be no stable state with a positive proportion p_{D1} as both p_{D1} and p_{D0} would continue to grow faster than their cooperation counterparts.

References

1. Isaac, M. R., McCue, K. F. & Plott, C. R. Public goods provision in an experimental environment. *J. Public Econ.* **26,** 51–74 (1985).
2. Isaac, M. & Walker, J. Group size effects in public goods provision: The voluntary contributions mechanism. *Q. J. Econ.* **103,** 179–199 (1988).

3. Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, Cambridge, 1990).
4. Olson, M. *The logic of collective action* (Harvard University Press, Harvard, 1965).
5. Hardin, G. The tragedy of the commons. *Science* **162,** 1243–1248 (1968).
6. Ostrom, E. Coping with tragedies of the commons. *Annu. Rev. Polit. Sci.* **2,** 493–535 (1999).
7. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
8. Hammerstein, P. editor. *Genetic and Cultural Evolution of Cooperation* (MIT Press, Cambridge, 2003).
9. Henrich, J. & Henrich, N. Culture, evolution and the puzzle of human cooperation. *Cog. Sys. Res.* **7,** 220–245 (2006).
10. Rand, D. A. & Nowak, M. A. Human cooperation. *Trends Cog. Sci.* **17,** 413–425 (2013).
11. Nowak, M. A. & Highfield, R. *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed* (Free Press, New York, 2011).
12. Von Neumann, J. Zur theorie der gesellschaftsspiele. *Mathematische Annalen* **100,** 295–320 (1928). English Translation Fin Tucker, A. W. & R. D. Luce, ed., Contributions to the Theory of Games IV, Annals of Mathematics Studies 40, 1959.
13. Von Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behaviour* (Princeton University Press, Princeton, 1944).
14. Nash, J. Non-cooperative games. *Ann. Math* **54,** 286–295 (1951).
15. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314,** 1560–1563 (2006).
16. West, S. A., El Mouden, C. & Gardner, A. Sixteen common misconceptions about the evolution of cooperation in humans. *Evol. Human Behav* **32,** 231–262 (2011).
17. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415,** 137–140 (2002).
18. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425,** 785–791 (2003).
19. Bowles, S. & Gintis, H. *A Cooperative Species: Human Reciprocity and Its Evolution* (Princeton University Press, Princeton, 2011).
20. Friedman, J. W. A non-cooperative equilibrium for supergames. *Rev. Econ. Stud.* **38,** 1–12 (1971).
21. Rubinstein, A. Equilibrium in supergames with the overtaking criterion. *J. Econ. Theor.* **21,** 1–9 (1979).
22. Kreps, D. M. & Wilson, R. Sequential Equilibria. *Econometrica* **50,** 863–894, July (1982).
23. Andreoni, J. Why free ride? strategies and learning in public goods experiments. *J. Public Econ.* **37,** 291–304 (1988).
24. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393,** 573–577 (1998).
25. Trivers, R. L. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46,** 35–57 (1971).
26. Sugden, R. *The Economics of Rights, Co-operation and Welfare* (Basil Blackwell, Oxford, 1986).
27. Alexander, R. D. *The Biology of Moral Systems* (Aldinee de Gruyter, New York, 1987).
28. Wedekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **288,** 850–852 (2000).
29. Milinski, M., Semmann, D., Bakker, T. C. M. & Krambeck, H.-J. Cooperation through indirect reciprocity: image scoring or standing strategy? *Proc. R. Soc. Lond. B* **268,** 2495–2501 (2001).
30. Bolton, G. E., Katok, E. & Ockenfels, A. Cooperation among strangers with limited information about reputation. *J. Public Econ.* **89,** 1457–1468 (2005).
31. Seinen, I. & Schram, A. Social status and group norms: Indirect reciprocity in a repeated helping experiment. *Eur. Econ. Rev.* **50,** 581–602 (2006).
32. Engelmann, D. & Steiner, J. Indirect reciprocity and strategic reputation building in an experimental helping game. *Games Econ. Behav* **67,** 399–407 (2009).
33. Nowak, M. A. & Sigmund, K. The dynamics of indirect reciprocity. *J. Theor. Biol.* **191,** 561–574 (1998).
34. Panchanathan, K. & Boyd, R. A tale of two defectors: the importance of standing for evolution of indirect reciprocity. *J. Theor. Biol.* **224,** 115–126 (2003).
35. Perc, M., Gómez-Gardeñes, J., Szolnoki, A., Flora, Y. & Moreno, L. M. Evolutionary dynamics of group interactions on structured populations: a review. *J. R. Soc. Interface* **10,** 20120997 (2013).
36. Sigmund, K. Punish or perish? retaliation and collaboration among humans. *Trends Ecol. Evol.* **22,** 593–600 (2007).
37. Rand, D. G. The promise of mechanical turk: how online labor markets can help theorists run behavioral experiments. *J. Theor. Biol.* **299,** 172–179 (2012).
38. Rand, D. G., Nowak, M. A., Fowler, J. H. & Christakis, N. A. Static network structure can stabilize human cooperation. *Proc. Natl. Acad. Sci. USA* **111,** 17093–17098 (2014).
39. Binder, K. & Hermann, D. K. *Monte Carlo Simulations in Statistical Physics*. Springer, Heidelberg, (1988).
40. Szabó, G. & Fáth, G. Evolutionary games on graphs. *Phys. Rep.* **446,** 97–216 (2007).

Acknowledgments

Author Contributions

H.H.N., M.P., A.S. and D.H. designed and performed the research as well as wrote the paper.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Nax, H. H. *et al.* Stability of cooperation under image scoring in group interactions. *Sci. Rep.* **5,** 12145; doi: 10.1038/srep12145 (2015).