Imitative Trust versus Indirect Reciprocity

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

[1]Kellogg School of Management and NICO, Northwestern University, Evanston, Illinois, USA, 60208

[2]CABDyN Complexity Centre, Oxford University, UK, OX1 1HP

[3]Oxford Centre for Integrated Systems Biology, Oxford University, UK, OX1 3QU

[4]Centre for Mathematical Biology, Oxford University, UK, OX1 3LB

[5]Institute for Science, Innovation, and Society, Saïd Business School, Oxford University, UK, OX1 1HP

August 31, 2009

Abstract

Indirect reciprocity, a key concept in evolutionary game theory, provides a mechanism that allows cooperation to emerge in a population of self-regarding individuals even when repeated interactions between pairs of actors are unlikely. Recent empirical evidence, in particular laboratory experiments, show that humans typically follow complex assessment strategies involving both reciprocity and imitation when making cooperative decisions. However, we currently have no systematic understanding of how indirect reciprocity and imitation mechanisms interact, and how they jointly affect the emergence of cooperation and the allocation of resources. 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 trust score, which tracks whether actors have been the recipients of cooperation in the past. We show that as long as trust scores are not used more than reputation scores, a population has a high likelihood of evolving into cooperative strategies. Surprisingly, we find that when information about an actor's reputation is hardly available, trusting the action of third parties towards her does favor a fairer allocation of resources compared to random-trusting and share-alike mechanisms. This might provide a new perspective to understand the use of imitation and reciprocity mechanisms in populations of cooperating social actors.

Summary

We study the emergence of cooperation and the fair allocation of resources in a population, where information about someone's reputation is frequently unavailable and actors might be influenced by the cooperative behavior of others. We investigate, using an evolutionary game theory approach, the consequences of the joint strategy of imitation and indirect reciprocity mechanisms, and why imitation is a recurrent mechanism in human behavior given its potential negative effects on the distribution of resources in a population. This theoretical advance in evolutionary game theory is also motivated by recent empirical evidence, in particular laboratory experiments, which show that humans typically follow com-

plex assessment strategies involving both reciprocity and imitation when making cooperative decisions. Comparing imitation to random-trusting and share-alike mechanisms, our findings reveal that trusting the actions of third parties towards an actor might be better alternative strategy to indirect reciprocity when information about that actor's reputation is unavailable. This suggests that imitative trust might be, in fact, an adaptive mechanism in populations of cooperating social actors.

1 Introduction

The evolution of cooperative behavior in biological and human populations has been shown to rely critically on different forms of reciprocity [1–5]. 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 [6, 7], and assign a reputation to the individual guided by the principle: if I scratch your back, someone else will scratch mine [8]. 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 especially if all individuals have access to the reputation scores of other individuals

[9–11].

However, when information about the past record of other individuals is unavailable or unreliable, laboratory experiments [12–18] and simulation models [13, 19–21] 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 experimental work suggests that imitation mechanisms may have evolved along with the cooperative behavior of individuals [22]. The trustworthiness is 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 [16, 17, 23, 24]. 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 [22, 25].

Although reliance on imitative 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 [23, 26–28]. 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. 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. The theoretical challenge that our paper addresses is to understand how imitation and indirect reciprocity mechanisms interact when information about an actor's reputation is frequently unavailable, and why imitation is a recurrent mechanism in human behavior given its potential negative effects on the distribution of resources in a population.

2 The model

In our imitation-reciprocity (IR) model, we consider individuals faced with a social dilemma [1, 8, 29, 30], who follow complex assessment strategies involving both reciprocity and imitation mechanisms [22, 25]. Here, the joint payoff is maximized by the cooperation between a donor and a recipient, but the donor's individual payoff is higher if she defects. 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 [9], 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. For 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 recipeint 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 addition, to take into account the important effects of memory constraints, bias in judgments or implementation errors [31–35], we introduce a fixed level of uncertainty into donor-decision making processes. This is implemented by allowing donors to randomly change their decision with a small, fixed probability q [10]. Here we consider that one out of ten times a donor can make an implementation or decision error ($q = 0.1$). Smaller values of q generate similar results (see Methods for a detailed description of the IR model).

3 Simulation results and discussion

3.1 The evolution of cooperation

In common with previous image-scoring models, we found that the IR model subject to any p value generates evolutionary stable strategies (ESSs) after a small number of generations. Note that an ESS is reached when the whole population adopts a common strategy without further change [36]. Figure 1 shows the diversity of ESSs (lines) in the population across a number of generations for different levels of trust usage p (colors).

In our model, the population moves from a purely cooperative behavior at $p = 0$ (i.e. pure indirect reciprocity) to a non-cooperative one at $p = 1$ (i.e. pure imitation). This reflects the fact that imitative trust is not a self-sustaining cooperative mechanism. However, we observed that the emergence of cooperation generated by indirect reciprocity is surprisingly robust to imitation mechanisms. As illustrated in Figure 2 (solid line), we found that under ESS behavior the average proportion of cooperative interactions (i.e. the ratio between the number of times any donor cooperates and the number of times any one does not) is over 50% for most of the simulated p-values. Measuring the frequency of possible ESSs of our IR model, Figure 2 (dashed line) shows that the use of imitation p also increases the uncertainty of ESSs as measured by Shannon's entropy [37] given by $-\sum_i p(ESS_i)log(p(ESS_i))$, where $p(ESS_i)$ is the likelihood of occurrence of ESS i. Note that the maximum uncertainty of possible ESSs is reached at $p = 0.8$, which also corresponds to the point where the proportion of cooperative interactions drops to less than 50% (see Figure 2 solid line). This implies that the higher the use of imitation mechanisms, the larger the uncertainty of the system and the lower the proportion of cooperation observed in a population.

3.2 The allocation of resources

To examine the allocation of resources produced by the joint assessment strategy of imitation and indirect reciprocity, we measure the distribution of cooperation in the population generated by changing the parameter p. 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 utility of actor i, and n is the total number of actors in the population. Hence, by using the utility of actors u_i as a proxy for wealth, we can measure the correlation between ESSs and the allocation of resources in the population. Figure 3A shows that under pure indirect reciprocity, 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 suggests that populations that have completely lost their confidence in reputation attributes and instead cooperate following imitation mechanisms are highly prone to inequality effects.

To investigate whether trusting the behavior of others is a useful cooperative mechanism when information about a recipient's reputation is unavailable, we replace our imitation mechanism by two different trusting strategies. Our first alternative strategy 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 with probabilty $p' = 0.5$. For the second alternative strategy, we assume that donors follow a share-alike behavior, where they try to distribute benefits equally among all members in the population [38, 39]. Here, donors cooperate if the trustworthiness (i.e. previous granted cooperation) of the recipient is low and defect if the trustworthiness is high (see Methods for details). Figures 4A-B and 4C-D show that under high values of p, random-trusting processes and share-alike mechanisms increase the likelihood of non-cooperative strategies (top left corner) and inequality effects (dark shades) compared to imitation mechanisms (see Fig. 3B-C). This reveals that trusting the behavior of others is, in fact, a useful alternative mechanism to indirect reciprocity when donors do not have access to the reputation of potential recipients.

4 Conclusions

Here we have analyzed for the first time the effects that two different 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 behavior of others. We find that both the cooperative behavior and the fair allocation of resources decrease as the use of imitation mechanisms increases. However, we have found that as long as actors use imitation and indirect reciprocity mechanism equally, cooperative strategies dominate and resource inequalities are small. Surprisingly, we have found that trusting the behavior of others generates lower inequality effects than simple random-trusting processes and share-alike mechanisms. This suggests that imitative trust might be in fact an adaptive mechanism in populations of cooperating social actors.

Methods

IR Model. Specifically, we consider that n actors interact over a fixed lifetime through m randomly chosen pair-wise interactions, where they can play either the role of donors or recipients. If a donor i cooperates with a recipient j, they generate a higher payoff b, which in this case is entirely allocated to the recipient's utility u_j. Otherwise, if the donor does not cooperate, then she can only produce a lower payoff $c < b$, which this time increases her own utility u_i. 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 [9]. 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. 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 s_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 $S = R = 6$. At the end of its lifetime, the population is replaced by a new generation, where an old actor can transmit her assessment strategies T-R to a new actor, with a probability proportional to her own utility following the standard replicator equation [36]. All generations start with $t_i = r_i = u_i = 0$ for all actors i. Simulations were performed using conventional parameter values [9, 10]: $m = 500$, $n = 100$, and $c/b = 0.1$. 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 $r_j \leq R_i$ for trust and reputation respectively.

Acknowledgements

We thank Robin Dunbar, Chiu-Fan Lee, Mason Porter, Janet Smart and Brian Uzzi for useful comments on a previous draft. SS held a Postdoctoral Fellowship at the Oxford University Corporate Reputation Centre in conjunction with the CABDyN Complexity Centre. FR-T acknowledges the support of EPSRC grant EP/E056997/1.

References

[1] Ostrom E, Burger J, Field C B, Norgaard R B, Policansky, D (1999) Revisiting the Commons: Local Lessons, Global Challenges *Science* **284**, 278-282.

[2] Axtell R L (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 P (Ed.) (2003) *Genetic and Cultural Evolution of Cooperation* (MIT Press).

[5] Saavedra S, Reed-Tsochas F, Uzzi, B (2009) A simple model of bipartite cooperation for ecological and organizational networks *Nature* **457**, 463-466.

[6] Wedekind C, Milinski M (2000) Cooperation Through Image Scoring in Humans *Science* **288**, 850-852.

[7] 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.

[8] Alexander R D (1987) *The Biology of Moral Systems* (Aldine de Gruyter).

[9] Nowak M A, Sigmund K (1998) Evolution of indirect reciprocity by image scoring *Nature* **393**, 573-577.

[10] Leimar O, Hammerstein P (2001) Evolution of cooperation through indirect reciprocity *Proc. R. Soc. B* **268**, 745-753.

[11] Pacheco J M, Santos F C, Chalub F A C C (2006) Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity *PLoS Computational Biology* **12**, e178.

[12] Pingle M (1995) Imitation versus rationality: An experimental perspective on decision making *J. of Socio-Economics* **24**(2), 281-315.

[13] Bikhchandani S, Hirshleifer D, Welch I (1998) Learning from the Behavior of Others: Conformity, Fads, and Informational Cascade *The J. of Economic Perspectives* **12**(3), 151-170.

[14] Gergely G, Bekkering H, Király, I (2002) Rational imitation in preverbial infants *Nature* **415**, 755.

[15] Goldstein D G, Gigerenzer G (2002) Models of ecological rationality: The recognition heuristic *Psychological Rev.* **109**, 75-90.

[16] Cialdini R B, Goldstein N J (2004) Social influence: compliance and conformity *Annual Rev. Psych* **55**, 591-621.

[17] Salganik, M J, Dodds P S, Watts D J (2006) Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market *Science* **311**, 854-856.

[18] Tomasello M (2008) *Origins of Human Communication.* (MIT Press)

[19] Banerjee A V (2002) A simple model of herd behavior *The Quarterly J. of Economics* **107**(3), 797-817.

[20] Bolton G E, Katok E, Ockenfelds A (2005) Cooperation among strangers with limited information about reputation *J. of Public Economics* **89**, 1457-1468.

[21] Na J P, Volken H, Pestelacci E, Tomassini M (2009) Conformity hinders the evolution of cooperation on scale-free networks *Phys. Rev. E* **80**, art. no. 016110.

[22] Fowler J H, Christakis N A (2009) Cooperative Behaviour Cascades in Human Social Networks arXiV:0908.3497.

[23] Merton R K (1968) The Matthew Effect in Science *Science* **159**, 56-63.

[24] Podolny J M (2005) *Sociological Signals: A Sociological Study of Market Competition* (Princeton Univ. Press).

[25] Ostrom E, Walker J (2003) *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research* (Russell Sage Foundation Publications).

[26] Zipf G K (1942) The unity of nature, least-action and natural social science *Sociometry* **5**, 48-62.

[27] de Solla Price D J (1976) A general theory of bibliometric and other cumulative advantage processes *Journal of the American Society for Information Science* **27**, 292-306.

[28] Ijiri Y, Simon H A (1977) . *Skew Distributions and the Sizes of Business Firms* (North-Holland).

[29] Bowles S (2008) Policies Designed for Self-Interested Citizens May Undermine The Moral Sentiments: Evidence from Economic Experiments *Science* **320**, 1605-1609.

[30] Gintis H, Bowles S, Boyd R T, Fehr E (Ed.) (2005) *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (MIT Press).

[31] Milinski M, Wedekind C (1998) Working memory constrains human cooperation in the Prisoners Dilemma. *Proc. Natl Acad. Sci. USA* **995**, 13755-13758.

[32] Ohtsuki H (2004) Reactive strategies in indirect reciprocity. *J. Theor. Biol.* **227**, 299314.

[33] Brandt H, Sigmund K (2006) The good, the bad and the discriminator-errors in direct and indirect reciprocity. *J. Theor. Biol.* **239**, 183194.

[34] Sommerfeld R D, Krambeck H-J, 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.

[35] Rankin D J, Eggimann F (2009) The evolution of judgement bias in indirect reciprocity *Proc. R. Soc. B* **276**, 1339-1345.

[36] Hofbauer J, Sigmund K (1987) *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press).

[37] Shannon C E (1948) Mathematical theory of communication *Bell System Technical J.* **27**, 379-423.

[38] Fehr E, Bernhard H, Rockenbach B (2008) Egalitarianism in young children *Nature* **454**, 1079-1084.

[39] Tomasello M, Warneken F (2008) Share and share alike *Nature* **454**, 1057-1058.

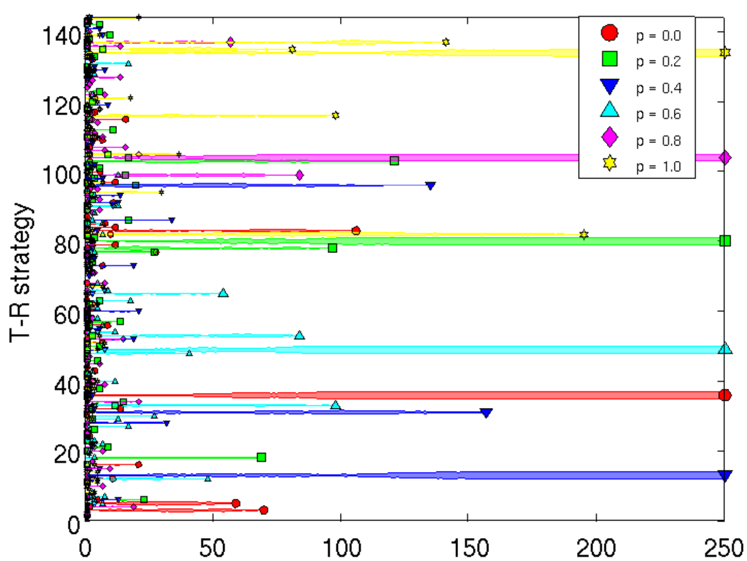

Figure 1: Evolutionary stable strategies. The figure shows the evolution of ESSs for a single simulation over different levels of trust p (markers/colors) for $n = 10^2$ members. The horizontal axis shows the number of generations m and the vertical axis presents the 144 different T-R strategies. Strategies 1-26 belong to $T < 0$ and $R < 0$, strategies 27-52 belong to $S < 0$ and $R > 0$, strategies 53-88 belong to $T > 0$ and $R < 0$ and strategies $T > 0$ and $R > 0$. Note that at the beginning of simulations ($m = 0$), the population starts with a heterogeneous mix of strategies. However, after few generations, all agents in the population adopt a single strategy. The line width represents the number of individuals in each ESS.

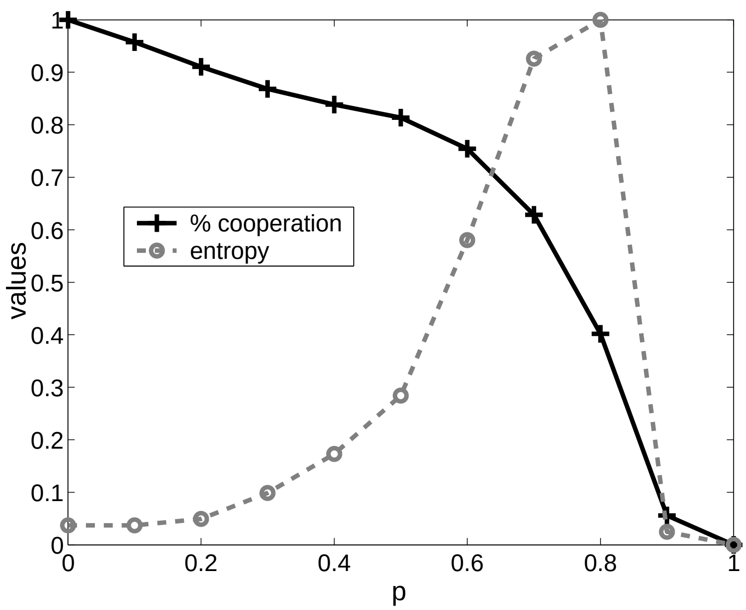

Figure 2: Cooperation and Information entropy. The solid line shows the correlation between the use of imitation mechanisms p and the proportion of cooperative interactions (without counting decision errors q make at the individual level) observed over 10^5 different simulation runs under ESS behavior. The dashed gray line shows the correlation between p and the uncertainty of information measured by Shannon's entropy (Shannon 1948) over the ESS space. The entropy is normalized to fall between 0, lowest observed value, and 1, maximum observed value. Note that the uncertainty of ESSs reaches a maximum at $p = 0.8$, and then declines up to the lowest level where only non-cooperative ESSs dominate.

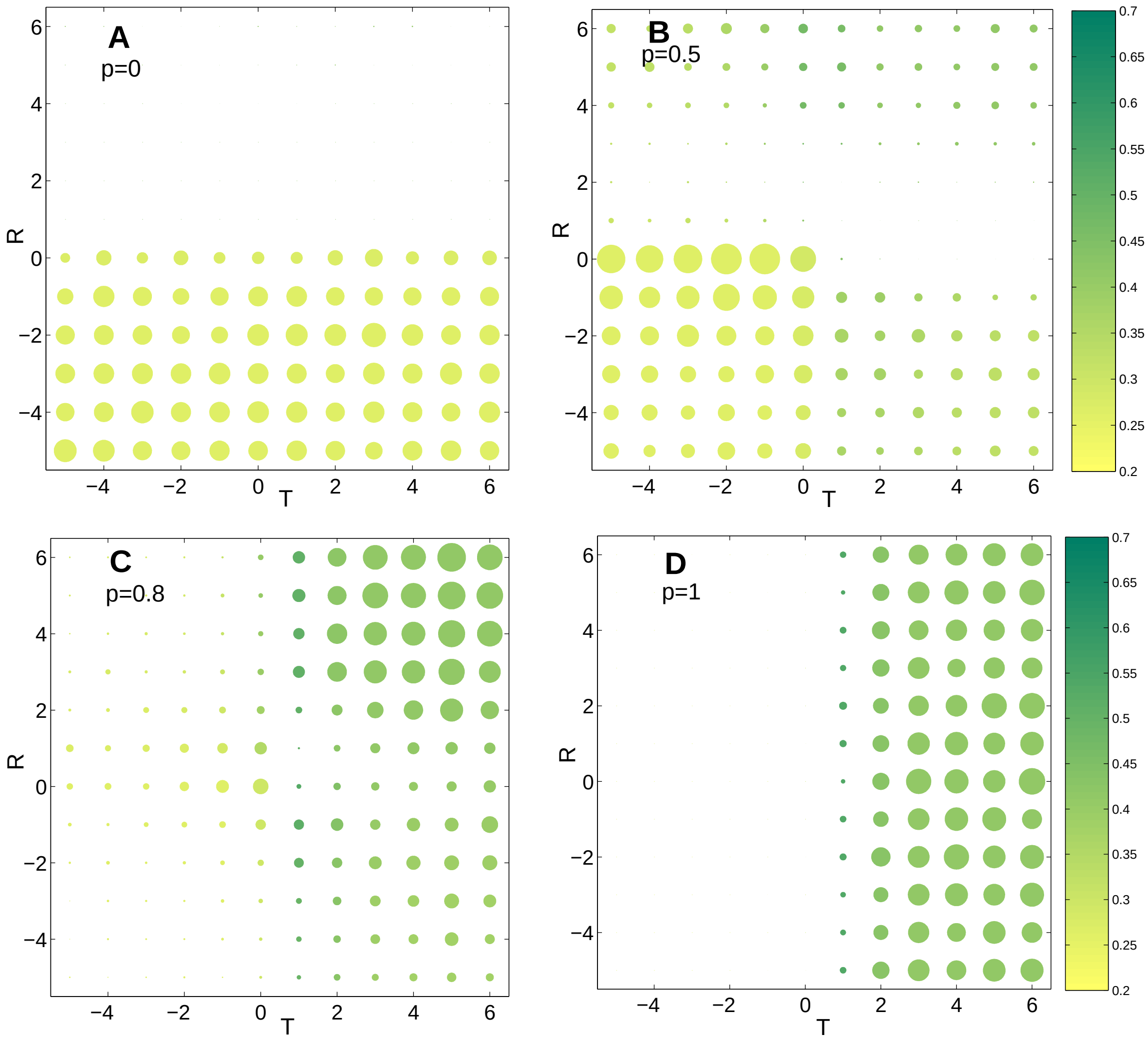

Figure 3: Gini coefficients and evolutionary stable strategies. Panels **A-D** show the correlation between Gini coefficients (shades) and the frequency of ESSs (circles) for $p = 0$, $p = 0.5$, $p = 0.8$ and $p = 1$ respectively. Gini coefficients and frequencies are reported as the average over 10^5 different simulation runs under ESSs. The frequency of occurrence for each ESS is proportional to the area of the circles. Note the presence of cooperative strategies ($T \leq 0$ and $R \leq 0$) for all level of p except $p = 1$.

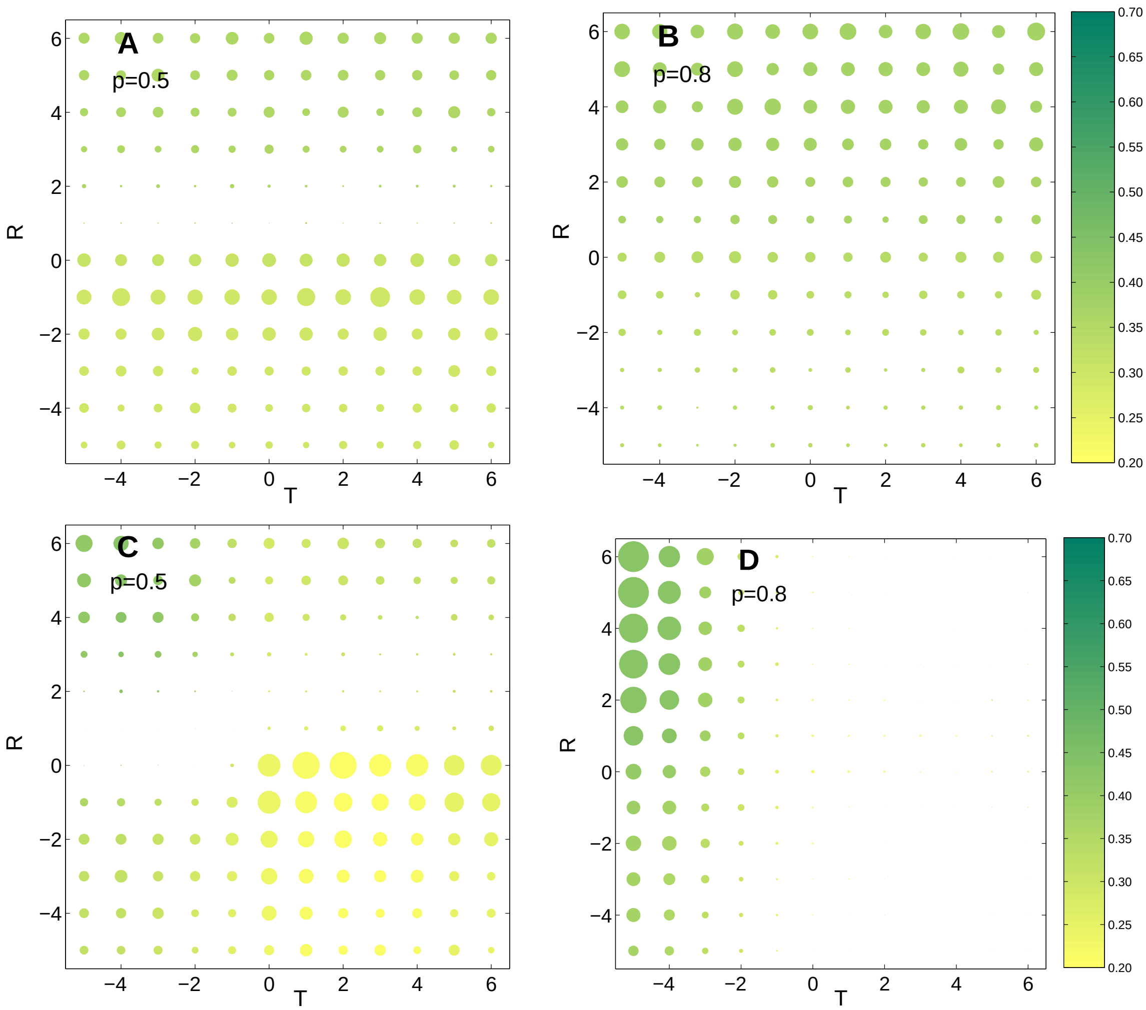

Figure 4: Alternative cooperative mechanisms to imitation. For random-trusting processes and share-alike mechanisms respectively, panels **A-B** and **C-D** show the correlation between Gini coefficients (shades) and the frequency of ESSs (circles) for $p = 0.5$, $p = 0.8$. Gini coefficients and frequencies are reported as the average over 10^5 different simulation runs under ESSs. The frequency of occurrence for each ESS is proportional to the area of the circles. Note the increase in Gini coefficients (dark shades) and ocurrence of non-cooperative strategies (top left corner) compared to those generated by imitation mechanism and indirect reciprocity as shown in Figures 3B-C.