

Evolution of Egalitarian Punishment

Kohei Tamura¹*, Rieko C. Morita¹, Yasuo Ihara¹

¹ Department of Biological Sciences, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

*Author for correspondence (tamura@biol.s.u-tokyo.ac.jp)

Punishment has been deemed as a key to solve the puzzle of how cooperation evolved. Recent studies have suggested that altruistic punishment may be motivated by preference for social equality (egalitarian punishment). Here we construct individual-based models to investigate the effectiveness of egalitarian punishment in promoting cooperation. Based on computational experiments in a meta-population model, we first show that egalitarian punishment is as effective as classic punishment, a form of punishment executed upon directly observing others' strategies. We then use a scale-free network model to show that egalitarian punishment can be effective even when heterogeneity in interactions among individuals is incorporated. Finally, we show that generosity in punishment can affect co-evolution of egalitarian punishment and cooperation.

Keywords

cooperation, egalitarian motives, punishment

Introduction

The evolution of cooperation is one of the greatest puzzles in both biological and social sciences. Theoretical and empirical studies have shown that cooperation is promoted if punishment of defectors is allowed at punisher's cost (e.g., Fehr & Gächter, 2002; Sigmund, 2007). However, what motivates such altruistic punishment is yet to be fully understood.

Recent studies have suggested that punishment may partially be motivated by preference for social equality. Dawes, Fowler, Johnson, McElreath, and Smirnov (2007) have reported that participants in a laboratory game were willing to pay a cost in order to reduce an above-average earner's income, even in the absence of any cooperative behavior to be reinforced. A connection between such egalitarian motives and altruistic punishment has been indicated by another experiment: participants who cared about equality were most likely to punish free-riders in public goods games (Johnson, Dawes, Fowler, McElreath, & Smirnov, 2009). In addition, Scheuring's (2010) simulation has shown that punishment by egalitarian motives

(hereafter, egalitarian punishment) can evolve when individuals cannot directly observe others' strategies. In other words, the possibility has been raised that egalitarian punishment may have played an important role in the evolution of cooperation.

We suspect, as did Scheuring (2010), that inferring others' strategies in humans is often difficult. If correct, punishing others on the basis of their strategies (hereafter, classic punishment) seems very improbable.

In this paper, we consider such a situation, and examine whether or not egalitarian punishment can facilitate the evolution of cooperation. We first explore whether or not egalitarian punishment can perform equally well in a setting where classic punishment is known to promote high levels of cooperation. More specifically, we substitute egalitarian punishment for classic punishment in a model developed by Boyd, Gintis, Bowles, and Richerson (2003), in which individuals are divided into groups and a public goods game with punishment is played within each group. In this model, interactions occur within a closed group and the number of interactions is constant, thus difference in payoffs directly reflects the difference in strategies. Hence, it can be anticipated that egalitarian and classic punishment would be directed at roughly the same individuals.

On the other hand, when interactions are more heterogeneous, the connection between payoff and strategy becomes less obvious. As a result, egalitarian punishers may punish a cooperator who simply has more cooperative partners than others. Taking this into account, we use scale-free networks to model interactions among individuals in the second part of our analysis. In contrast to the meta-population model utilized by Boyd et al. (2003), a scale-free network allows heterogeneous interactions. Human social networks often exhibit characteristics of a scale-free network and this fact might have played a key role in the evolution of human cooperation (Masuda, 2007; Santos, Pacheco, & Lenaerts, 2006).

We also investigate how generosity in egalitarian punishment may influence the evolution of cooperation. Generosity has been defined variously in the literature (e.g., Kurokawa, Wakano, & Ihara, 2010; Nowak & Sigmund, 1992). Here we define it as the maximum amount of above-average gain to be tolerated without being punished.

By conducting computational experiments, we ask: (i) how effective egalitarian punishment is in a meta-population model, (ii) whether egalitarian punishment can promote the evolution of cooperation on scale-free networks, and (iii) in

what ways generosity may affect the evolution of cooperation.

Model 1

We followed Boyd et al.'s (2003) model except for the condition of punishment. Consider a population divided into 128 groups of n individuals playing two-stage public goods games. There are three strategies: cooperators, egalitarians and defectors, whose frequencies are denoted by x , y and z , respectively. In the first stage, individuals choose to cooperate or to defect. Cooperation incurs a cost c to produce a benefit b that is shared equally among group members. Defection incurs no costs and produces no benefits. Cooperators and egalitarians cooperate with probability $1-e$ and otherwise defect. Defectors always defect. In the second stage, an egalitarian punishes individuals who obtained larger payoffs than the group mean. The cost of being punished is p/n and the cost of imposing punishment is q/n . The expected fitness of cooperators, egalitarians and defectors are $w+b(x+y)-c$, $w+b(x+y)-c-qz$ and $w+b(x+y)-py$, respectively, where w represents the baseline fitness. After the second stage, each individual is paired with a randomly chosen partner either from its own group with probability $1-m$ or a partner from another randomly chosen group with probability m . Individual i encountering individual j imitates the latter's strategy with probability $W_j/(W_j+W_i)$, where W_k is the fitness of individual k . Groups are paired at random and inter-group conflict occurs with probability e . The probability that group i defeats group j is $1/2[1+(z_j-z_i)]$, where z_k is the frequency of defectors in group k . Finally, mutation occurs within each individual with probability μ which results in switch of strategy into one of the two other strategies. Initially, one group consists of only egalitarians and the other 127 groups are composed of only defectors.

Simulations were run for 2000 time periods. Long-term average frequencies of cooperators and egalitarians were obtained by averaging over the last 1000 time periods. Figure 1 shows the average frequency of cooperation over 1000 runs.

Our result shown in Figure 1 is qualitatively identical to that reported by Boyd et al.'s (2003): inter-group competition and punishment promote the evolution of cooperation even when the size of group is large. We concluded, therefore, that egalitarian punishment is as effective as classic punishment in promoting high levels of cooperation.

Model 2

Here we consider a scale-free network with a population of $N = 5000$ playing two-stage Prisoner's Dilemma game. We used the Barabási-Albert algorithm (Barabási & Albert, 1999) to generate scale-free networks. In this algorithm, networks expand continuously by the addition of new nodes, and new nodes attach preferentially to the existing

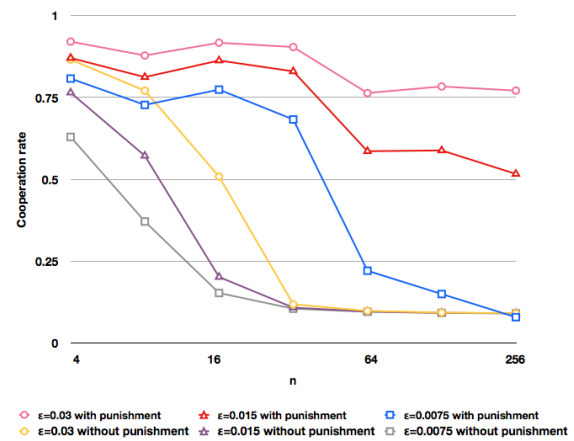


Figure 1. Egalitarian punishment is as effective as classic punishment in promoting cooperation. When punishment is possible, $p = 0.8$ and $q = 0.2$. Other parameter values are: $b = 0$, $c = 0.2$, $m = 0.01$, $e = 0.02$, $\mu = 0.01$ and $w = 1.0$.

nodes that are already well connected. The average number of neighbors was set to four.

Following the models proposed by Nakamaru and Iwasa (2006) and Rand, Armao, Nakamaru, and Ohtsuki (2010), individuals are divided into two types, altruistic (A) or selfish (S) according to the behavior in the first stage. They are also dichotomized into either egalitarian punisher (P) or non-punisher (N) for their behaviors in the second stage. Thus, four combinations of types are possible: AP, AN, SP and SN. In the first stage, an altruist from each pair of individuals connected in the network pays a cost c to increase the fitness of its neighbor by benefit b . The total payoff for each individual is calculated by summing the payoffs over all of its interactions. In the second stage, each egalitarian punisher compares payoffs obtained in the first stage to those of its neighbors, including itself, and punishes those who have gained above average. Suppose, for example, an egalitarian punisher i is connected to individuals j and k , whose payoffs are P_i , P_j and P_k , respectively. Egalitarian punisher i punishes individual j if $P_j > (P_i + P_j + P_k)/3 + g$, where g represents generosity. The cost of being punished is denoted by p and the cost of punishing by q .

As an update rule, we adopted the Fermi rule: each individual randomly selects one of its neighbors and copies its strategy with probability $1/[1+\exp(-\alpha(P_{\text{opp}}-P_{\text{self}}))]$, where α represents the intensity of selection and is set to 10 in what follows. Mutation was set to occur within each trait after imitation with probability $\mu = 10^{-5}$. The initial frequency of each strategy was 0.25 on average.

Simulations were run for 6000 time periods. Long-term average frequency of each strategy was obtained by averaging over the last 1000 time periods. Results shown in Figures 2 to 5 are based on the averages over 1000 runs (5 runs for each of 200 different scale-free networks).

With or without punishment, when the benefit of cooperation, b , is above a threshold, the cooperation rate increases and reaches a plateau as b

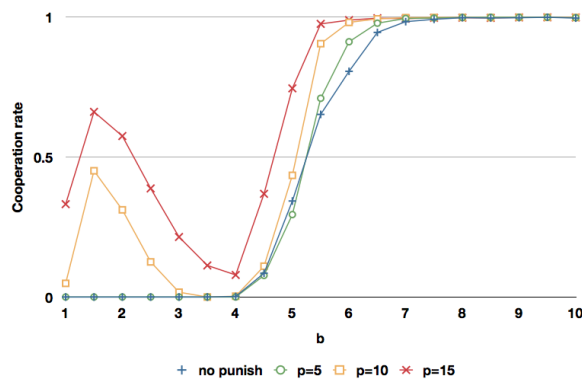


Figure 2. Egalitarian punishment can promote cooperation when the benefit of cooperation is small. Parameter values are: $c = 1$ and $q = 1$ (when punishment is possible).

increases (Figure 2). In the presence of egalitarian punishment, the cooperation rate is also heightened within a range of b that is below the threshold, which is observed as a peak of the cooperation rate in Figure 2. At the peak, most of the altruists are punishers, while most of them are non-punishers at the plateau (Figure 3). This suggests that cooperation is enhanced by two different mechanisms: spatial structure at the plateau and egalitarian punishment at the peak.

Generosity has both positive and negative impacts on the evolution of cooperation: it reduces the risk of altruistic individuals being punished but increases the chance of selfish individuals evading punishment. Figure 4 shows that the optimal level of generosity, which achieves the highest cooperation rate, depends on b . It also suggests that generosity shifts the peak of cooperation rate toward smaller values of b .

One way of measuring the efficiency of egalitarian punishment for a given generosity is to calculate the difference between the proportion of altruistic individuals being punished and that of selfish individuals being punished (Figure 5). The value of b that gives the maximum extra punishment on selfish individuals shifts toward

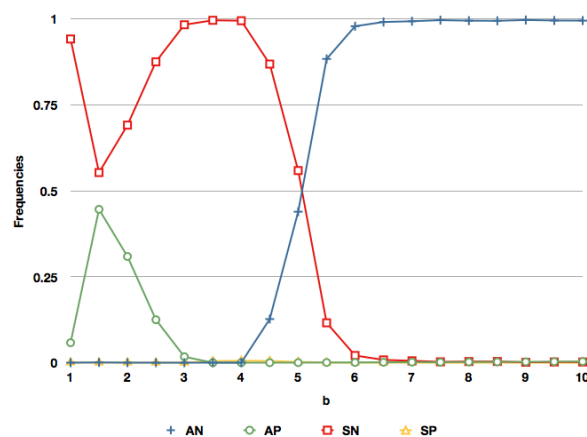


Figure 3. Most of altruistic individuals are punishers at the peak and non-punishers at the plateau. Parameter values are: $c = 1$, $p = 10$ and $q = 1$.

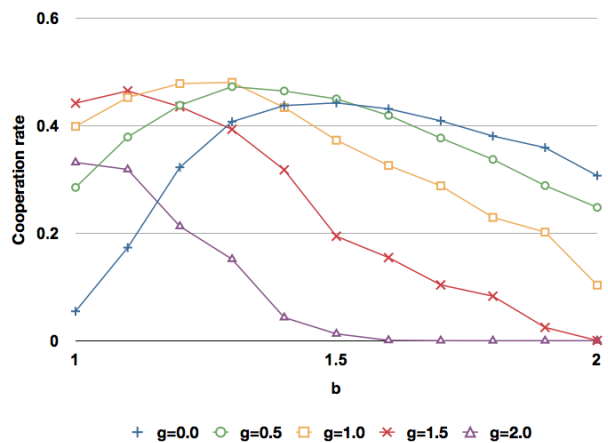


Figure 4. Generosity affects the cooperation rate. Parameter values are: $c = 1$, $p = 10$ and $q = 1$.

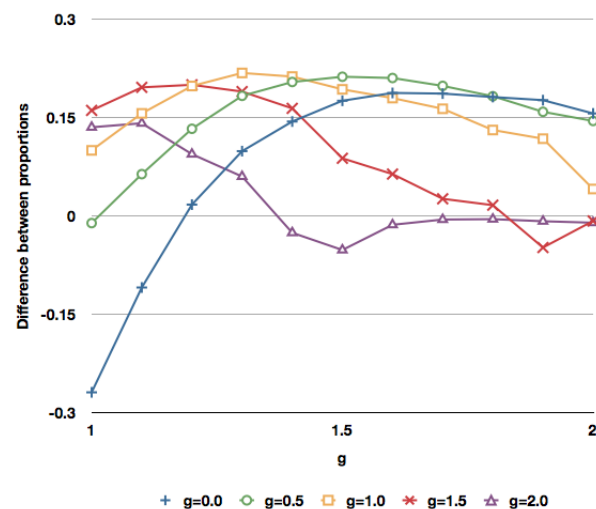


Figure 5. Generosity affects extra punishment on selfish individuals relative to that on altruistic individuals. The vertical axis represents the difference between the proportion of selfish individuals being punished and that of altruistic individuals being punished. Parameter values are: $c = 1$, $p = 10$ and $q = 1$.

smaller values of b with increasing generosity, which is concordant with the peak shift in cooperation rate (Figure 4).

Discussion

The current paper explored the co-evolution of cooperation and egalitarian punishment through computational experiments in a meta-population model and a scale-free network model. We first demonstrated that in the meta-population model, egalitarian punishment is as effective as classic punishment in terms of promoting cooperation. This result is intuitive because in the meta-population model in which interactions are homogenous among individuals, both egalitarian and classic punishments work roughly the same way.

We then examined whether or not egalitarian punishment could promote cooperation in scale-free networks, where heterogeneity in the interactions

was taken into consideration. Although spatial structure alone could promote cooperation if the benefit of cooperation (b) was sufficiently large, egalitarian punishment was able to give rise to cooperation when b was too small for the spatial structure to do so. When the average cooperation rate was plotted against b , the effect of spatial structure was observed as a plateau that is reached as b becomes sufficiently large. The effect of egalitarian punishment was represented as a peak, indicating that egalitarian punishment was effective within a certain range of b . Outside of this range, egalitarian punishment seemed ineffective either because the punishment was excessively directed at altruistic individuals, or the benefit gained by selfish individuals was too large to be offset by the cost of being punished.

We also showed that generosity could affect the co-evolution of cooperation and egalitarian punishment. Generosity could promote cooperation by reducing punishment on altruistic individuals, though excessive generosity also fostered selfish individuals. The optimal level of generosity that achieves the highest cooperation rate depended on b . In addition, the peak of cooperation rate shifted toward smaller values of b as generosity increased. A major, but not necessarily exclusive, factor contributing to the peak shift is the change in relative probabilities with which altruistic and selfish individuals are punished.

A limitation of the present study would be that we considered only one update rule, because the outcomes of evolutionary games in spatial structures are known to vary according to the details of update rules (Nakamaru & Iwasa, 2005, 2006). Further research is needed to examine the robustness of our results with different update rules and a wider region of parameter space.

We do not claim that egalitarian preference is the only motive that nurtures effective punishment. Nevertheless, our results suggest that egalitarian motives can be an important factor for the evolution of cooperation.

References

- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509-512. (doi:10.1126/science.286.5439.509)
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the USA*, 100, 3531-3535. (doi:10.1073/pnas.0630443100)
- Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, 446, 794-796. (doi:10.1038/nature05651)
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140. (doi:10.1038/415137a)
- Johnson, T., Dawes, C. T., Fowler, J. H., McElreath, R., & Smirnov, O., (2009). The role of

- egalitarian motives in altruistic punishment. *Economics Letters*, 102, 192-194. (doi:10.1016/j.econlet.2009.01.003)
- Kurokawa, S., Wakano, J. Y., & Ihara, Y. (2010). Generous cooperators can outperform non-generous cooperators when replacing a population of defectors. *Theoretical Population Biology*, 77, 257-262. (doi:10.1016/j.tpb.2010.03.002)
- Masuda, N. (2007). Participation costs dismiss the advantage of heterogeneous networks in evolution of cooperation. *Proceedings of the Royal Society B*, 274, 1815-1821. (doi:10.1098/rspb.2007.0294)
- Nakamaru, M., & Iwasa, Y. (2005). The evolution of altruism by costly punishment in lattice-structured populations: score-dependent viability versus score-dependent fertility. *Evolutionary Ecology Research*, 7, 853-870.
- Nakamaru, M., & Iwasa, Y. (2006). The coevolution of altruism and punishment: Role of the selfish punisher. *Journal of Theoretical Biology*, 240, 475-488. (doi:10.1016/j.jtbi.2005.10.011)
- Nowak, M. A., & Sigmund, K. (1992). Tit for tat in heterogeneous populations. *Nature*, 355, 250-253. (doi:10.1038/355250a0)
- Rand, D. G., Armao, J. J., IV., Nakamaru, M., & Ohtsuki, H. (2010) Anti-social punishment can prevent the co-evolution of punishment and cooperation. *Journal of Theoretical Biology*, 265, 624-632. (doi:10.1016/j.jtbi.2010.06.010)
- Santos, F. C., Pacheco, J. M., & Lenaerts, T. (2006). Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences of the USA*, 103, 3490-3494. (doi:10.1073/pnas.0508201103)
- Scheuring, I. (2010). Egalitarian motive in punishing defectors. *Journal of Theoretical Biology*, 264, 1293-1295. (doi:10.1016/j.jtbi.2010.02.047)
- Sigmund, K. (2007). Punish or perish? Retaliation and collaboration among humans. *Trends in Ecology & Evolution*, 22, 593-600. (doi:10.1016/j.tree.2007.06.012)