Klaus Jaffe and Luis Zaballa (2010)

Co-Operative Punishment Cements Social Cohesion

Journal of Artificial Societies and Social Simulation **13** (3) 4
<http://jasss.soc.surrey.ac.uk/13/3/4.html>

Received: 04-May-2009 Accepted: 20-Nov-2009 Published: 30-Jun-2010

Abstract

Most current attempts to explain the evolution—through individual selection—of pro-social behavior (i.e. behavior that favors the group) that allows for cohesive societies among non related individuals, focus on altruistic punishment as its evolutionary driving force. The main theoretical problem facing this line of research is that in the exercise of altruistic punishment the benefits of punishment are enjoyed collectively while its costs are borne individually. We propose that social cohesion might be achieved by a form of punishment, widely practiced among humans and animals forming bands and engaging in mob beatings, which we call co-operative punishment. This kind of punishment is contingent upon—not independent from—the concurrent participation of other actors. Its costs can be divided among group members in the same way as its benefits are, and it will be favoured by evolution as long as the benefits exceed the costs. We show with computer simulations that co-operative punishment is an evolutionary stable strategy that performs better in evolutionary terms than non-cooperative punishment, and demonstrate the evolvability and sustainability of pro-social behavior in an environment where not necessarily all individuals participate in co-operative punishment. Co-operative punishment together with pro-social behavior produces a self reinforcing system that allows the emergence of a 'Darwinian Leviathan' that strengthens social institutions.

Keywords: Altruism, Cooperation, Social, Prosocial, Cohesion, Evolution, Punishment, Retribution

Introduction

1.1 Why most people stop at a read traffic light and pay taxes? Why people normally help a tourist in finding directions? Why a soldier offers his life to help advance his countries interest? Why people donate blood, give money to charities, or spend time helping others? These are fundamental questions we need to answer if we want to understand human society. Recent research has shown that mutualism and reciprocity among non related individuals emerges if cooperation provides synergies which will benefit all actors (Trivers 1971, Axelrod 1984, Nowak and Sigmund 1998, Panchanathan and Boyd 2004, Ohtsuki et al. 2006). The feature here that tilts the balance in favor of pro-social behavior is the amount of social synergy achieved by social cooperation (Jaffe 2001, 2002). That is, social cooperation can be achieved and maintained even among non related individuals when:

$$S = \sum b$$
$$c < p \cdot S \text{ and } \alpha = S/c$$

where *p* is the proportion of all the future benefits triggered by the pro-social behavior received by the actor; *S* is the total sum of all benefits triggered by the act; and α is the social synergy or the ratio of benefits/costs gained through cooperation. This inequality describes the total benefits acquired by society through social investments. Simulations (Jaffe 2001, 2002) and empirical evidence (Osborn and Jaffe 1997) suggest that when $\alpha >> 1$, social behavior is evolutionary stable.

1.2 When social synergy can not explain social behavior (i.e. when $\alpha \sim 1$ and individuals are not related) pro-social behavior might be stabilized trough punishment of non-pro-social individuals or free-riders. The occurrence of altruistic punishment, through which individuals punish other individuals for failing to act pro-socially, increase the costs of free-riding, and thereby promotes pro-social behavior. The problem is that altruistic punishment is also costly, so rational individuals would, again, be more inclined to let others assume the costs, while still enjoying the fruits of the resulting pro-social behavior.

1.3 Many propositions have been made that aim to explain how to overcome this shortcoming (for example; Bowles and Gintis 2006; Boyd 2006; Fehr 2000; Gintis 2003; Gintis et al. 2003; Sachs et al. 2004), searching for a possible evolution of pro-social behavior as a result of individual selection, but this line of research has not yet produced an accepted explanation of the evolutionary dynamics leading to pro-social altruism. The essential problem here is that each group member is tempted to act pro-socially in order to reap the fruits of the social welfare resulting from the concurrent pro-social efforts of other group members, but more strongly tempted to spare the individual costs of pro-social altruism while still enjoying those fruits, so the predictable outcome is the disappearance, or nonappearance, of pro-social behavior.

1.4 To tackle this problem researchers have introduced the possibility of what they call altruistic punishment, through which individuals punish other individuals for failing to act pro-socially, in order to increase the costs of free-riding, and thereby improve and promote the option of pro-social behavior. The problem is that altruistic punishment is also costly, so rational individuals would, again, be more inclined to let others assume the costs, while still enjoying the fruits of the resulting pro-social altruism. This is known as the 'second order public goods problem', which may in turn be addressed by introducing the possibility of second order altruistic punishment, that is, punishment for those who fail to punish free-riders. But since this form of punishment is also costly, the same problem would arise, meaning that a 'third order punishment' would be needed, and so on. So the problem would be indefinitely displaced to a higher order of punishment, but never actually solved.

1.5 This is the essential problem faced today by theorists of the evolution of pro-social behavior through altruistic punishment (Fehr 2000, Bowles et al. 2003, Boyd et al. 2003; Gintis 2000; Hauert et al. 2007; Sánchez and Cuesta 2005). The core of the problem, we believe, is the assumption that the punishment required to enforce pro-social altruism has to be applied individually-without possibly coordinating efforts with other group members-as in a prisoners' dilemma situation. We fail to see why members of a social group could not apply punishment co-operatively-instead of individually-which would enable them to distribute the costs of punishment evenly among all group members. And if such costs can, in fact, be distributed among group members, the cost to each individual is minimized and the theoretical problem of the understanding the evolutionary dynamics of pro-social behavior may actually be solved (Zaballa 2006).

1.6 Co-operative punishment is a fundamental part of human society -as exemplified by human bands, Indian tribes, slum mobsters, the police, law enforcement, taxation and most modern social institutions- and might occur in other animal societies although we could not find any published evidence for this. Another route to co-operative punishment is mobbing. Although as described originally it is aimed at predators, it is used to harass co-operatively something that represents a threat to them, mobbing against conspecifics would classify as co-operative punishment. Unfortunately experimental evidence for behaviors like mobbing or other cooperative strategies to punish intra-specific free-riders among animal societies is very scarce or totally absent. (Conradt and List 2009)

1.7 In order to gain a better insight into the evolutionary dynamics of co-operative punishment in

stabilizing pro-social behavior, we simulated the interaction of different forms of cooperation and free-riding, using agent based computer simulations.

Computer Simulations

2.1 We modeled the evolution of a virtual population of 50 hunter-gatherers by means of a simulation program called Sociodynamica (2000), previously used to model economic aspects of altruistic co-operation (Jaffe 2001), altruistic punishment (Jaffe 2004), and the role of shame in stabilizing co-operation (Jaffe 2006). Simulations with larger populations — 500 and 2,000 agents — produced similar results and can be performed by downloading the software (Sociodynamica 2000). The model emulates widely used experimental economics game in which each member of a group is provided an endowment, b (food in Sociodynamica) that increases every time-step in 3 units. These units can be kept for future consumption and reproduction or can be invested in a public good. The combined investment in the public good is multiplied by a factor, _, and distributed equally to everyone in the group. The total payoff of each individual (the proportion of the endowment kept for oneself plus one's share of the public good) is related to fitness as excess food is used to produce offspring. In the present set of experiments, $\alpha = 1$. The alpha parameter is the degree of synergy the cooperation produces. Increasing " α" will increase the odds for cooperative strategies to invade the population, as larger values for α increases the incentive for cooperation and reduces the incentives for defection. (Jaffe 2002, 2004). Thus, $\alpha = 1$ is a very stringent condition for cooperators to survive.

2.2 The accumulated wealth- fitness (W) of either cooperators (co) or free-riders (fr) is:

$$W_{co} = n_{co}\, b + s - c$$

$$W_{fr} = n_{fr}\, b + s - p$$

where:

n_{co} = total number of cooperators

n_{fr} = total number of free-riders

b = amount of resources received trough feeding (constant)

c = cost of cooperation (constant)

p = cost of punishment · probability of being punished if a free-rider

The benefit received through social cooperation (s) is defined as:

$$S = (n_{co} \cdot c \cdot \alpha - \sum p') / (n_{co} + n_{fr})$$

where:

α = synergy achieved trough social cooperation

p' = cost to punish the captured free-riders

2.3 Described in plain English, individuals either foraged alone and did not cooperate with anybody (free-rider), or agreed to join part of their hunting and gathering efforts in order to form every season (time step) a common pile of food (contributor). Each individual collected 3 food units a season, and if socially mined, contributed to the common pile with one unit. The resulting common pile was distributed evenly among all group members, independently of their contribution. After eating 0.5 food units per season, each member assigns his remaining wealth to self-reproduction, at a cost of 2 units per clone. The cloned offspring suffered a 10% mutation rate. A lifetime consisted of only 10 seasons, and random death kept the population at 50 individuals.

2.4 At the beginning of a simulation, half of the actors were contributors, while the other half were free-riders. Later on, gene frequencies varied according to reproductive success. When simulating the emergence of sociality, the simulation began with all individuals as free-riders and after 3 time-steps, offspring was allowed to eventually mutate to be a contributor.

2.5 The simulation scenarios explored were as follows:

No Punishment (NP): All individuals may contribute to the common pile. Enforcement relies entirely on individual's good will, with no monitoring and punishment.

Altruistic Punishment (AP): To tackle the free-riders problem, individuals are allowed to enforce the social contract by punishing those they encounter and that failed to contribute to the common pile. Punishment involves a detraction of food units twice the value of the withheld contribution. Since free-riders will presumably resist being punished, altruistic punishers will also incur certain costs, and since all individuals are presumed equally hurtful, those costs will equal those of punishment itself.

Cooperative punishment (CP): As in AP but punishment is performed by the group so that the cost of every punishment act is subtracted from the common pile. Punishers and punished individuals paid the same cost as in AP, but group members apply punishment co-operatively, meaning that the costs of punishing free-riders are not assumed by a few freelance punishers, but distributed equally among all members of the society.

2.6 The summary of the different variables used in the simulations is given in Table 1.

Table 1: Variables defining the rules of the game in the simulations

Society	Defined by the use of the social contribution *C*. Societies simulated were: No Punishment (*NP*), Altruistic Punishment (*AP*) or Co-operative (Collective) Punishment (*CP*)
C	Contribution. Was paid as a proportion of the wealth accumulated by the agent. All agents with *s*=1 paid their contribution.
Y	Cost to the punisher. In the present simulations Y = C
K	Cost of the fine extracted to the punished agents
E	Efficiency in punishing free-riders (agents with *s* = 0). This efficiency is given as the percentage of free-riders punished. In NP, E = 0.
P	Percentage of pro-social agents

2.7 The sequence of programmed actions was:

```
  Create initial population distributing individual parameters randomly
  Distribute resources in space
 100 Make individual walk in Brownian motion
  Individuals collect resources
  Monitor encounter of individuals
  Determine commercial exchange between individuals
  Determine altruistic exchanges
  Determine punitive exchanges
  Rearrange matrix
  Pay taxes
  Select survivors
  New agents are created
  Introduce mutations in new agents
  Goto 100 unless maximum number of time steps has been reached
  Produce output and statistics
```

For more details see Sociodynamica 2000.

Results

3.1 Simulation results showed clearly that Cooperative Punishment is the strategy most likely to maintain high proportions of contributing individuals in the population. This advantage is evidenced in

a range of border conditions. The most important condition in achieving a stable population with a high proportion of contributors to the common good is the relation between the cost of the punishment (K) and the cost incurred by the punisher (Y) (Figure 1)

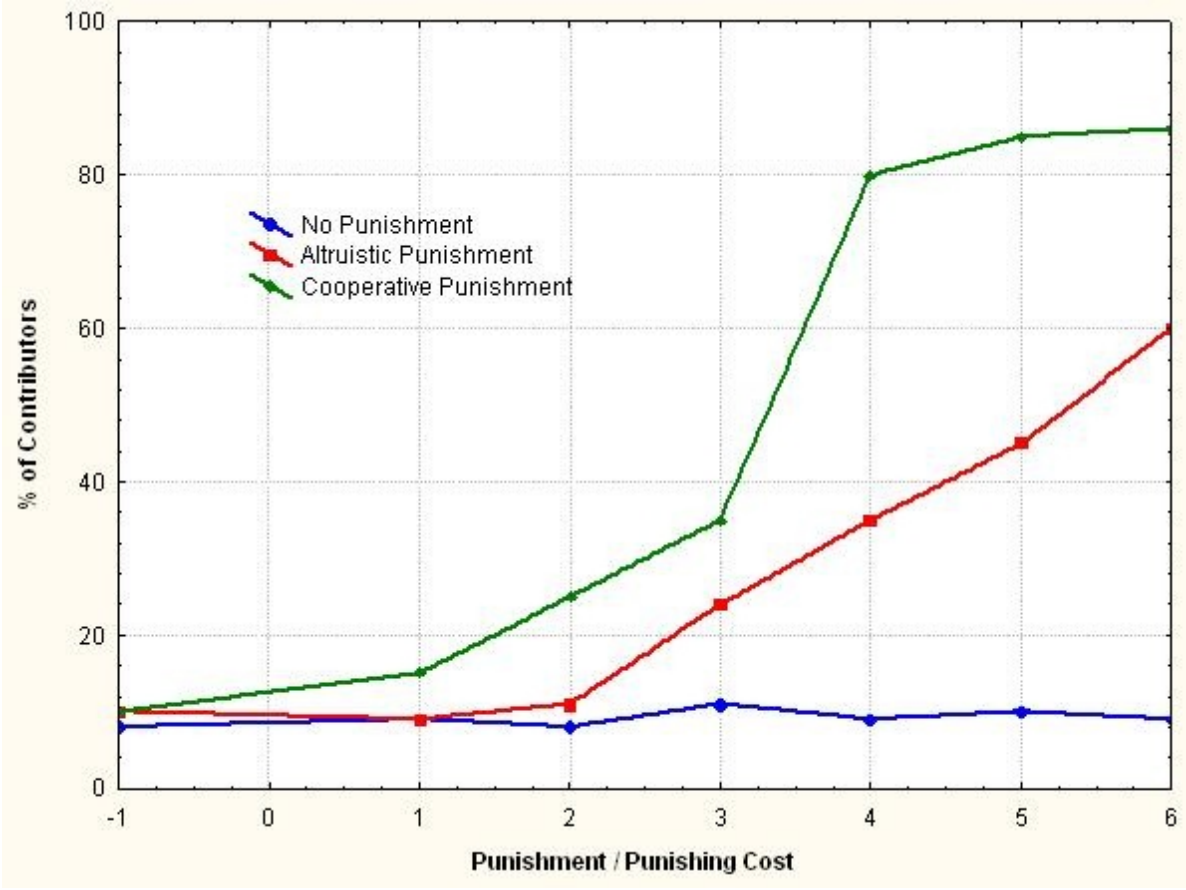

Figure 1. Summary of simulation results showing the average of the Percentage of Contributors in the population when simulating three different societies with No Punishment, societies were Altruistic Punishment was possible and societies were Cooperative Punishment was enforced. The x axis shows the different ratios between the costs of the Punishment (or cost to the punished: K) and the cost to the punishers (Y) used in the simulations. The efficiency of reaching free riders for punishment E = 60%.

3.2 Figure 2 shows that when punishment costs and cost to punish are very low or very large, the incentive of punishment to maintain individuals to contribute to the common good dissipates, even with Cooperative Punishment, as it either has no punishment effect (K=0) or is too expensive to be financed by the common pool (Y>4)

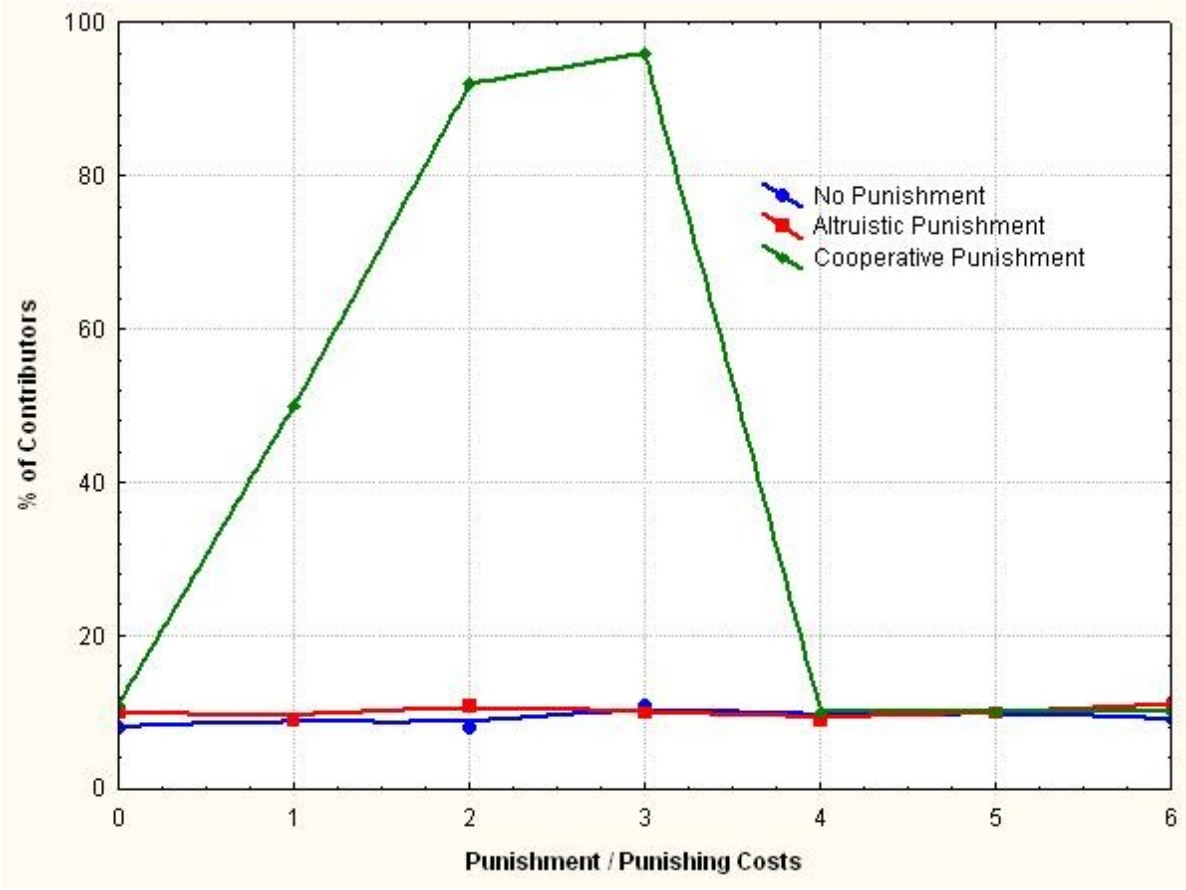

Figure 2. Results of simulations showing the average percentage of contributors reached when simulating three different societies with contribution $C = 1$, at different Punishment Costs: $K = Y$. $E = 100$. Data points represent the average of 200 simulations run to time step 200.

3.3 The efficiency in catching free-riders for punishment also affects the likelihood of maintaining a majority of contributors. Figure 3 shows that when less than 60% of free-riders are punished, punishment, even of the Cooperative Punishment kind, is not able to maintain contributions to the common pile.

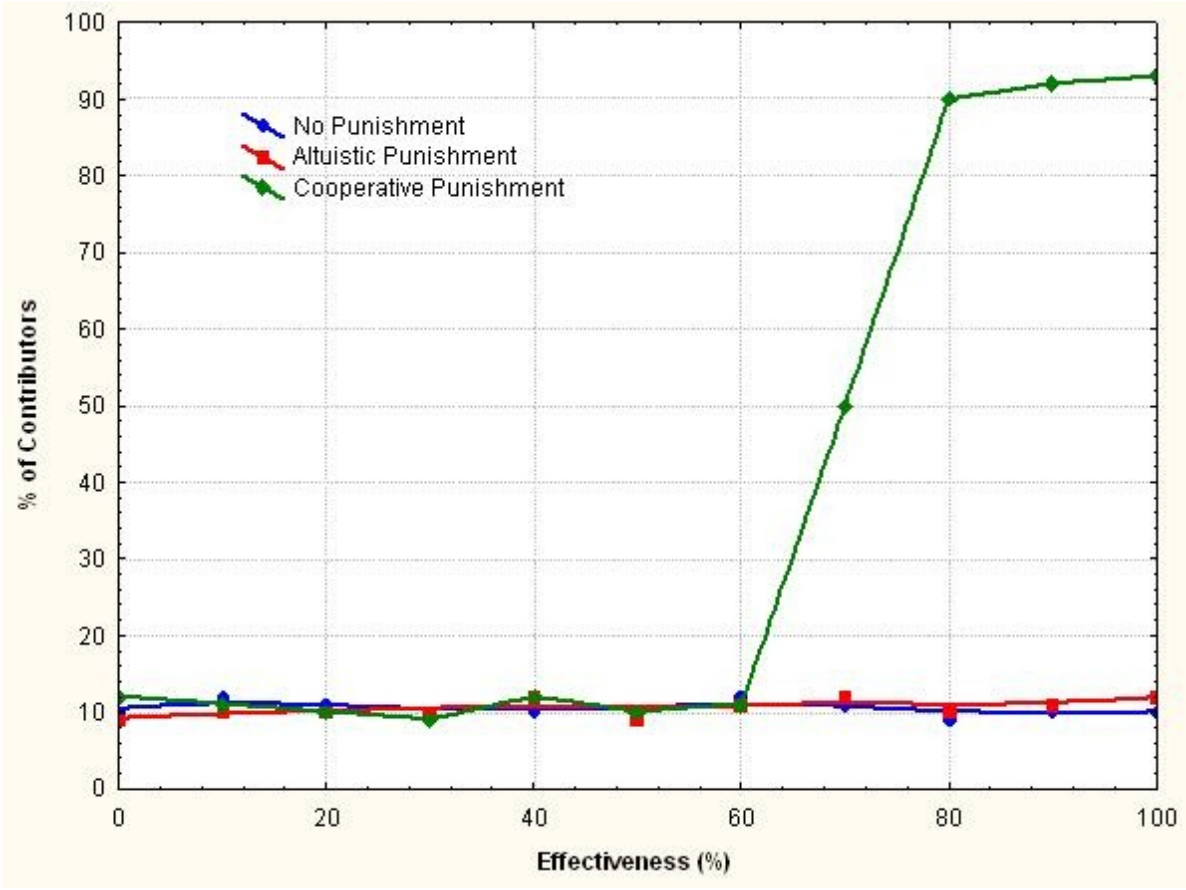

Figure 3. : Simulation results showing the average values for the Percentage of Contributors from simulations of the three different societies, where $C = 1$, $K = Y = 2$, The effectiveness of catching free-riders: E varied from 0 to 100, as indicated in the horizontal axis. Data points represent the average of 200 simulations run to time step 200.

3.4 We run also simulations exploring the likelihood for pro-social behavior to emerge. In these simulations all individuals in the initial populations were free-riders and after 3 time steps, mutant individuals that behaved as contributors were allowed to appear and simulations were run for 1000 time steps. The results showed that for conditions represented in Figure 1 where K/Y ratios favored the maintenance of pro-social behavior, contributors could invade a population of free-riders if it was driven by Cooperative Punishment but not if Altruistic Punishment was the form of punishment. Even for K/Y ratios larger than 6 were Altruistic Punishment strategies unlikely to foment pro-social behavior (see Table 2).

Table 2: Percentage of runs (%) that showed a majority of pro-social individuals after 1000 time steps when the initial population was 100% free-riders for societies using Altruistic Punishment or Cooperative Punishment. Using the same runs, the average number of pro-social individuals from 100 simulation runs.

K / Y	AP (%)	CP (%)	AP (average)	CP (average)
1	0	100	10.5	72
7	2	100	35	99.6

Conditions Y = 1, K = 1 or 7, E = 100, C = 1

3.5 Running simulations where in addition to Cooperative Punishment individuals could implement Altruistic Punishment showed that Altruistic Punishment tended to disappear but increased the chances for the establishment of pro-social behavior.

Conclusions

4.1 We showed that co-operative punishment is a much stronger stabilizer of pro-social behavior than altruistic punishment. This advantage is achieved mainly by improving the efficiency and the economics of punishing free-riders. The simulations showed that reducing the K/Y ratio of punishment increases the evolutionary stability of pro-social behaviors. The simulations explored simple idealized conditions and we are aware that other features may favor the establishment and maintenance of social punishment, such as:

- Co-operative punishment may reduce the costs of punishment as a consequence of the synergy that typically results from co-operation. For example, when various individuals punish someone co-operatively, resistance may be expected to fall dramatically reducing the cost for punishing and thus increasing the ratio: cost to punished / cost to punisher
- Co-operative punishment may increase the effectiveness of punishment as a result of the combined capacities of many society members in monitoring individual behavior, making it possible to detect infractions in a way that freelance punishers could never match.
- Social enforcement of rules is less subject to forces affecting the individual, such as a lack of immediacy, or immediately available resources for punishment, etc., and thus more efficient by itself, irrespective of all advantages cited above
- Co-operative punishment may involve additional costs in terms of observations, evaluations, and discussions required to reach an agreement. In constituted societies punishment costs may actually lie for the most part in these necessary proceedings rather than in the execution of punishment itself, thus reinforcing its power to exert a consistent selective pressure leading to the evolution of pro-social behaviors.

4.2 Co-operative punishment might be common among animals (Ratnieks and Wenseleers 2008) or it might be that the cognitive requirements for co-operative punishment to work make it more likely to be found among human society (Boehm 1999). Studies on higher primates show that some advanced features of complex pro-social behavior can also be found in non-human animals (Gomes et al. 2009). In any case, humans enforce pro social structures by co-operative punishment following the same basic pattern as mob-beatings, for society members carefully avoid assuming the costs of punishment individually, but press for public resolutions that divide the costs of punishment among all society members. One way to achieve this is reputation through moral gossip, by which individuals make public their private knowledge of other people's antisocial behavior until there is a consensus to apply some form of punishment. If after a series of antisocial acts people agree, for instance, that the offender should be ostracized-a common punishment in band societies that in practice may amount to death penalty-the costs of such punishment, which consist mainly of loosing the co-operative capacities of the offender, are practically nil. This kind of cooperation might be especially important in keeping religious groups together (Jaffe and Zabala 2009). Another way to socialize the costs of punishment is to appoint punishers (police among humans; individuals specialized in tackling social corruption among social insects (Zweden et al 2007 for example) and compensate them with public resources-the common pile of food in our modeled society-so that the costs of punishment are ultimately borne by all society members, whether they actually participate in punishment or not.

4.3 If co-operation solves the second order public goods problem, why would it not solve the first order public goods problem (Eldakar and Wilson 2008 for example)? What our simulations show is that co-operative punishment together with pro-social behavior produces a self reinforcing system that allows the emergence of a social Leviathan. Thomas Hobbes (1651) described the origin of human society as the result of a social agreement to build a powerful Leviathan (the State) capable of punishing antisocial behavior on behalf of the people. As shown here, enforcing co-operative punishment require no additional behavior to those already present in the cooperative web of behaviors that are enforced by the punishment. In our simulations, the same pile of food that assures the common good is used to pay for punishments. In the light of our findings, the emergence of cooperation becomes less of a puzzle and we might rather start asking why cooperation is not ubiquitous in animal and human society (Herman et al. 2008). Our results suggest that the lack of strong institutions, or corruption inside institutions responsible for cooperative enforcement of rules,

may play an important part in explaining the failure of many a human society. This seems to be an accepted view by many mainstream economists (World Bank 2004). Much remains to be explored regarding co-operative punishments in other animal societies.

References

AXELROD R (1984) *The Evolution of Cooperation*. (New York: Basic Books).

BOEHM C (1999) *Hierarchy in the forest: Egalitarianism and the evolution of human altruism* . Harvard University Press, Cambridge (Massachusetts), 304 pp.

BOWLES S, Fehr E and Gintis H (2003) Strong reciprocity may evolve with or without group selection. *Theoretical Primatology Project Newsletter* December issue.

BOWLES S, Gintis H (2006) The Evolutionary Basis of Collective Action. Pp. 951-967 in *The Oxford Handbook of Political Economy* edited by Barry R. Weingast & Donald A. Wittman. (Oxford: Oxford University Press).

BOYD R (2006) The Puzzle of Human Sociality. *Science*, 314: 1553. [doi:10.1126/science.1136841]

BOYD R, Gintis H, Bowles S and Richerson, P J (2003) The Evolution of altruistic punishment. *Proceedings of the National Academy of Sciences* 100 6: 3531-3535 [doi:10.1073/pnas.0630443100]

CONRADT L and List C. (2009) Group decisions in humans and animals: a survey. *Philosophical Transaction of the Royal Socciety Lond B Biol Sci*. 364:719-42.

ELDAKAR O T and Wilson D S (2008) Selfishness as second-order altruism. *Proceedings of the National Academy of Science* 105: 6982 -6986 [doi:10.1073/pnas.0712173105]

FEHR E (2000) Cooperation and punishment. *American Economic Review* 90: 980-994.

FEHR E and Gächter S (2002) Altruistic Punishment in Humans. *Nature* 415:137-140 [doi:10.1038/415137a]

GÄCHTER S, Renner E and Sefton M (2008) The Long-Run Benefits of Punishment, *Science* 322: 1510.

GINTIS H (2000) Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology* 206:169-179. [doi:10.1006/jtbi.2000.2111]

GINTIS H (2003) The Hitchhikers Guide to Altruism: Genes and Culture, and the Internalization of Norms. *Journal of Theoretical Biology* , 220: 407-418. [doi:10.1006/jtbi.2003.3104]

GINTIS H, Bowles S, Boyd R and Fehr E (2003) Explaining altruistic behavior in humans. *Evolution and Human Behavior,* 24:153-172. [doi:10.1016/S1090-5138(02)00157-5]

GOMES C M, Mundry R and Boesch C (2009) Long-term reciprocation of grooming in wild West African chimpanzees. *Proceedings of Biological Science* . 276: 699-706 [doi:10.1098/rspb.2008.1324]

HAUERT Ch, Traulsen A, Brandt H, Nowak M A and Sigmund K (2007) Via freedom to coercion: the emergence of costly punishment. *Science* 316: 1905-1907.

HERMAN B, Thöni C and Gächter S (2008) Antisocial punishment across societies. *Science* 319: 1362-1367. [doi:10.1126/science.1153808]

HOBBES T 1651. *Leviathan, The Matter, Forme and Power of a Common Wealth Ecclesiasticall*

and Civil.

JAFFE K (2001) On the relative importance of Haplo-Diploidy, Assortative Mating and Social Synergy on the Evolutionary Emergence of Social Behavior. *Acta Biotheoretica* 49: 29-42. [doi:10.1023/A:1010229506863]

JAFFE K (2002) An Economic Analysis of Altruism: Who Benefits from Altruistic Acts? *Journal of Artificial Societies and Social Simulation* 5(3)3 http://jasss.soc.surrey.ac.uk/5/3/3.html

JAFFE K (2004) Altruism, altruistic punishment, and decentralized social investment," *Acta Biotheorica*, 52: 155-172. JAFFE K (2006) Simulations show that shame drives social cohesion" In: (Eds.) Sichman J. S. et al. IBERAMIA-SBIA (Berlin: Springer Verlag,) 88-97.

JAFFE K and Zaballa L (2009) Cooperative Punishment and Religion's Role in the Evolution of Pro-social Altruism. Chapter 13 in *The Biology of Religious Behavior: The Evolutionary Origins of Faith and Religion*. Edited by Jay R. Feierman. Praeger.

NOWAK M A and Sigmund, K (1998) Evolution of Indirect Reciprocity by Image Scoring. *Nature* 393: 573. [doi:10.1038/31225]

OHTSUKI H, Hauert C, Lieberman E and Nowak M A (2006) A Simple Rule for the Evolution of Cooperation on Graphs and Social Networks. *Nature* 441: 502-505. [doi:10.1038/nature04605]

OSBORN F and Jaffe K (1997) Cooperation vs. exploitation: interactions between Lycaenid (Lepidoptera : Lycaenidae) larvae and ants. *Journal of Research on the Lepidoptera* 34: 69-82.

PANCHANATHAN K and Boyd R (2004) Indirect Reciprocity can Stabilize Cooperation without the Second-Order Free Rider Problem. *Nature* 432, 499-502. [doi:10.1038/nature02978]

RATNIEKS FLW and Wenseleers T (2008) Altruism in insect society and beyond: voluntary or enforced? *Trends in Ecology and Evolution* 23: 45-52. [doi:10.1016/j.tree.2007.09.013]

SACHS J L, Mueller, U G, Wilcox T P and Bull, J J (2004) The Evolution of Cooperation. *Quarterly Review of Biology* 79: 135-160. [doi:10.1086/383541]

SÁNCHEZ A, and Cuesta J A (2005) Altruism may arise from individual selection. *Journal of Theoretical Biology* 235: 233-240.

SOCIODYNAMICA (2000) http://atta.labb.usb.ve/Klaus/Programas.htm

TRIVERS R (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology* 46: 35-57 [doi:10.1086/406755]

WORLD BANK (2004) *World Development Report 2004: Making Services Work for Poor People* . Oxford University Press.

ZABALLA L (2006) *Polis: A Natural History of Society* . (Kiev: Dukh i Litera, Mohyla Academy, National University of Kiev).

ZWEDEN J S, Fürst M A, Heinze J and d'Etorre P (2007). Specialization in policing behaviour among workers of the ants *Pachycondyla inversa*. *Proceedings of the Royal Society, London* B. 274: 1421-1428.