Complexity 14(2): 46-52, 2008.

Evolution of shame as an adaptation to pro-social punishment and its contribution to social cohesiveness

Klaus Jaffe
Universidad Simón Bolívar
Apartado 89000, Caracas 1080, Venezuela
kjaffe@usb.ve

Abstract: Feelings of shame are common among humans although shameless individuals do not seem to be handicapped in achieving social success in life. What then is the adaptive value of shame? How can shame have evolved? Here I simulate shame as the emotion that induces an increase in pro-social behavior after receiving social punishment. Simulations with the agent based model Sociodynamica show that shame is evolutionary stable in a context of individual selection, without the need for including group selection as an evolutionary force. The adaptive advantage of shame is based on the fact that it increases flexibility to the shameful individual, allowing it to act selfish if the probabilities of being punished are low and achieving a reduction in the costs of social punishment when frequent punishment is likely. The results show that shame, together with pro-social punishment and social cooperation, produce a fluctuating dynamics of social cooperation, achieving long periods where the populations stabilizes pro-social behavior interspersed with periods where selfish behavior predominates. This temporal stabilization of pro-social behavior might provide societies with sufficient time to built institutions that might sustainably stabilize pro-social behavior.

Introduction:

Understanding the evolution of cooperation is a challenge of modern sociobiology [1, 2, 3, 4, 5, 6, 7]. Group selection scenarios are very powerful in explaining the emergence of cooperation and societies [8], but the detailed conditions for group selection to work are not likely to occur in nature [9, 10] and thus, evolutionary explanations based on individual selection are deemed more robust than those based on group selection.

Regarding evolution based solely on individual selection we know that several evolutionary routes may lead to cooperation [11]. One set of routes is driven by kin selection through the emergence of altruistic behaviors favoring genetic related individuals and/or spiteful behaviors harming non-related individuals. The term altruism refers here strictly to behaviors where the actor incurs in a net cost to benefit another individual, and spite where the actor incurs in costs to harm another individual. Altruism and spite can be explained in terms of biological evolution, were gains and losses are measured in resources and reproductive potential, and were a behavior may eventually favor third parties. These interactions are explained by Hamilton's rule [12], which states that a behavior is evolutionary stable if the relatedness of the individual that profits from the altruistic act of the focal individual must be higher than the cost/benefit ratio this act imposes ($r > c/b$: r=genetic relatedness, c=costs, b=benefits).

Another set of behaviors that favor cooperation and sociality are independent of the genetic relatedness of individuals. In mutualistic behaviors or reciprocity, the balance of costs and benefits between the participants of the interaction are critical in determining the stability of cooperation.

In exploitative behaviors, the actor gains benefits at a cost to the receiver, which in dynamic terms mirrors altruism but with the label for actor and receiver reversed. Exploitative behaviors among non related individuals trigger competition and evolutionary arm races. Mutualism in contrast emerges if cooperation provides synergies which will benefit all actors [13, 14, 15, 16]. Behaviors that achieve social synergies are quite common [17] and are the main focus of microeconomics and business science. Several behavioral interactions, analyzed at small time windows, seem to be altruistic or exploitative, but seen through large time windows reveal occasional benefits to the donor, so that if the cost/benefit ratio is averaged over large time-spans and over several individuals they show to be advantageous to the average actor and thus may be favored by individual selection. These behaviors, also referred to as social investment, are evolutionary stable if investment is small compared to average future returns [18].

Several social institutions favor cooperation. One such institution is social punishment, often referred to as altruistic punishment, i.e. punishment to non-cooperative individuals at a cost to the punisher [19, 20, 21, 22, 23, 24, 25]. Social punishment imposes a cost to defectors of cooperation, thus making defection less likely. In human societies, social punishment often has no cost and triggers feelings of shame. Shame was proposed as a social cement by classical Greek philosophers such as Aristotle. More recently shame was shown experimentally to favor cooperation in real and virtual societies [2, 26, 27]. Here I explore the dynamic interactions between feelings of shame, mutualistic cooperation and social punishment.

Methods:

An overview of the model:

The computer model Sociodynamica, is an adaptation of the simulation model Biodynamica, designed to explore the evolution of behavior in artificial societies. The simulations aim at reproducing biological evolution more faithfully, including epistasis and pleiotropy. The present simulations aimed at mimicking behaviors relevant to primitive hominid communities, and may be also to advanced social primates. The model has been shown to be useful for clarifying concepts and pinpointing weaknesses in theories about economic and social behavior. It has been used to study the effect of altruism and altruistic punishment on the aggregate wealth in artificial societies [14, 25, 28,], the effect of division of labor on the economy [29], and the effect of shame on social cohesion [27]. It simulates the evolution of agents inhabiting a continuous two-dimensional toroidal world (500x400 pixels), which possess patches of food and minerals. Agents wander continuously with Brownian motion trough this world, each at its proper speed. Agent eats food in order to survive, which they collect from the food patches, if they happen to wander over it, or they get from other agents through barter and altruistic interactions. Possession of minerals by agents, which are also acquired by direct collection from the landscape or through barter, reduces the odds of being affected by catastrophes that occur randomly in time and space.

```
Prepare Simulation
    Create resources in the environment
    Create initial population of agents
Beginning of Simulation
    Produce Brownian motion for each agent
    Check for food and for neighbors
        Consume food if you are over it
        Interact with agent that is closer than 20 pixels
            Generous Donation
```

Social Punishment
Survive
Reproduce
Go to Beginning of Simulation

The virtual society has different levels of interaction.

1- Interactions with the environment: agents collected food and/or minerals.
2- A basic economy: barter of goods between agents.
3- Altruistic compassion: agents can donate excess food to starving agents.
4- A sense of justice: agents may punish stingy (non altruist) agents, with a cost to themselves.
5- Feeling of shame: shameful agents increase their altruistic donations after receiving punishment; whereas shameless individuals are not affected by punishment.
6- Social investment: individuals donated food but recovered their investment by increasing their value of minerals, which reduced their risk of suffering fatal catastrophes. This mimics the possibility that generosity buys solidarity, which might be decisive in critical moment.

A version of the model, including the exact programming code, can be downloaded from http://atta.labb.usb.ve/Klaus/klaus.htm and run in a Microsoft-Windows environment.

The virtual landscape:

The toroidal world was supplied with patches of land with food and others with minerals. Agents acquired a single unit (w_o) of the corresponding resource, each time they happened to land over one of these resources while walking randomly around, accumulating wealth, either as food (w_f) and/or as mineral wealth (w_m). The probability of arriving and resting over a patch of food or minerals was dependent on the speed and type of movement of the agent, defined by the variable DM in Table 1.

Both types of resources were replenished continuously. Each of them was concentrated in a different single patch and the total amount of resources was 400x400 w_o for food and 200x200 w_o for minerals. Each resource patch was distributed initially at random in the landscape but remained in the same place during the duration of each run.

Agents consumed food at a basal constant rate ($b = 0.5\ w_o$/time step). Agents with no food resources left ($w_f = 0$) perished. The wealth in food (w_f) of each agent changed each time step:

$$w_{ft} = -b\ w_{ft-1} + w_0$$

where $w_0 = 0$ if no resources are encountered; and

$$d\ w_f = -b\ dt$$

Similarly, agents encountering minerals acquired a single unit of the resource (w_o) each time they encountered it. Minerals never degraded.

Rules for survival:

Food was consumed by each agent at a constant rate (b) and a minimum amount was needed for reproducing. Possession of minerals reduced the odds of suffering a stochastic fatal catastrophe following an exponential function, so that larger amounts of minerals were needed to increase survival marginally.

Each agent's fitness was determined by two independent components: survival modulated by food (S_f) and minerals (S_m), so that to survive agent had to have positive values for

$S_f = w_f - w_r$ and $S_m = r^*(w_m/ (m^*D))$

The odds of survival modulated by food (S_f) were dependent on the amount of food accumulated (w_f) minus that used for reproduction (w_r), which in turn was dependent of the allele for fertility (*Fert* in Table 1); and survival modulated my minerals (S_m) was dependent randomly (r with values from 0 to 1) on the levels of accumulated wealth in minerals (w_m), the mean wealth in minerals accumulated by all agents (m), and a constant (D) representing dangers.

The wealth in minerals (w_m) was inversely related to the probability of suffering a fatal accident. Mineral wealth improved the odds of surviving external "catastrophes" that killed agents at random, each time step, and large amounts of w_m protected the agents against these catastrophes by reducing the probability of being affected by them. Agents with $w_m = 0$ could survive, though, with a lower probability. The odds of surviving fatal catastrophe is given by S_m. Agents did not survived an accident when: $w_m < \text{Rnd}(0\text{-}1) * D$

Reproduction

At the end of each time step, dead agents were replaced by new ones. This substitution process allowed maintaining the total number of agents in the population constant. Only the agents with food reserves had a significant chance to reproduce, producing the number of offspring according to the value of their fertility allele. That is agents with food reserves larger than 2 w_o, were selected at random to produce offspring, until the allowed maximum number of agents (1000) was reached. Offspring inherited all the alleles from their parent. Asexual reproduction was implemented as it provides for an appropriate simulated metaphor of real life [30, 31]. A 10 % of the genes of new agents suffered mutations that changed their allele randomly with another in the range allowed.

Genes characterizing the agents:

Each agent had a specific set of parameters or genes defining its behavior. Table 1 shows the parameters that defined the behavior of each agent and the range for values for each parameter (alleles). This system allowed to produce different types of agents with behaviors that can be operationally characterized true the effect on themselves and others. Compassionate pro-social altruist were agents with *ASt*=1 and Gen>0. Shameful agents were those with *Sha*>0, whereas shameless agents has Sha=0. Moralist punishers had *ASt*=1 and *PTh*>0. Business investors had *ASt* =2 and *Gen*>0. All intermediate combinations were allowed to occur.

Table 1: Genes characterizing the agents with their possible alleles (values).

DMo	Movement Speed	Distance in number of pixels agent moves every time step. Values vary from 0 to 20 pixels.
Fert	Fertility	Number of new agents produced when reproducing asexually. Values vary from 0 to 5 offspring.
ASt	Altruistic Strategy	Values of 0: agents do not engage in altruistic behavior or punishment. Values of 1 make agents donate food, not receiving any benefits from the altruistic act. Vales of 2 allow food donations to increases the amount of effective minerals possessed by the factor *Gain*, reducing the odds of succumbing to fatal catastrophes.
Gain	Gain	Factor by which the effective amount of minerals increases after a food donation when ASt = 2. Values for this factor vary from 0 to 10 so that the increase in minerals = amount of food donated**Gain*/5.
ATh	Altruistic Threshold	Times the altruistic agent has to have more food than the receiver in order to trigger a donation of food. Values vary from 0 to 10.Allele 0 simulates exploitation as poor agents transfer wealth to rich ones.
Gen	Generosity	Amount of food donated in the altruistic act. Values represent the deciles donated from 0 to10 (10 = 100%).
PTh	Punitive Threshold	Amount of generosity in the other agents below which altruistic punishment is triggered. Values vary from 0 to 5. Agents with *PTh*= 0 do not punish.
Cost	Cost of Punishment	Amount in food, punisher and punished agent pay. Values vary from 1 to 5 units (w_o).
Sha	Sensitivity to Punishment	Amount (in deciles) agents increased their generosity after receiving a punishment. Values vary from 0 to 5 deciles.

Social interactions:

Trading goods using barter: Agents moved in random directions each time step. Each time an agent met another at a distance smaller than 20 pixels, an exchange of wealth could occur. These could be of two different types: of food for minerals or of minerals for food, and occurred if the agent had an excess of one of the commodities i.e. w_f or $w_m > 2\ w_o$. All agents could collect food and/or minerals and could trade their goods at a ratio of one to one.

Altruistic generosity and egoistic cooperation: Donations of food occurred when the difference in food wealth (w_{f1}/w_{f2}) between the two agents was larger than the value given by their allele in *ATh*; then the richer agent transferred food to the less wealthy. If *ATh*=0, agent interchanged food regardless of the wealth of the other. The amount of food transferred depended on the generosity (*Gen*) of the donating agent, which varied initially among agents from 0 to 5 deciles of their wealth (w_f), i.e. 0 to 50 % of their wealth. Two types of cooperation were simulated: altruistic compassion (*ASt*=1) and egoistic (*ASt*=2) reciprocity. Altruistic compassion implied donations of food without receiving any reward other than helping stabilizing mutualistic cooperation of the population. Egoistic reciprocity implied that donations of food were rewarded with w_m increasing S_m, as an analogy to buying social capital, reputation, social insurance or goodwill, through altruistic acts. The value of the reward w_m was given by the gene *Gain*. If *ASt*=0 agents did not donate food.

Shame: Was defined as the drive that evokes a feeling of pain if one's behavior is disliked or rejected by others. It was operationalized by focusing on its social function, so that higher levels of shame increase the intensity of compliance with the social norm that elicited that feeling of shame. Agents were shamed by other agents acting as altruistic punishers (PTh>0), when the value of *Gen* in the contacted agents was lower than that of *PTh* of the punisher. The shamer imposed a fixed cost to the shamed agent and had to pay that same cost himself regulated by the alleles in the gene *Cost*. The intensity of the reaction of agents when shamed was coded by alleles in *Shame*, which regulated the degree shamed agents increased their generosity after having been shamed. This increase could vary from 0 to 5 deciles of w_f in *Gen*

Results

Figure 1 shows nine examples of simulations. Using color codes, the frequency in the population of each allele from 5 or 6 loci, can be monitored simultaneously. For example, in the simulation given in column 1, row 1, all organisms were inclined to provide generous donations (ASt=1), and all were shameless (Sha=0). After a few time steps, the organisms never engaging in punishment (Pun=0) and having the lowest possible generosity (Gen=0) prevailed. Adaptation reduced to cost of punishment to minimum levels at the beginning of the evolutionary process (Cost=0), but once most organisms were not punishing anymore, the alleles for cost drifted more randomly. Organisms adapted to maximize fertility (Fert=5). This last adaptation occurred in all simulations and will thus not be indicated further.

In the simulation given in column 1, row 2, a similar simulation was run but with costs for punishment held at a minimum (Cost=0) and organisms were allowed to have shame. Despite being able to evolve shame, most organisms remained shameless (Sha=0), but adaptation under these conditions did not drive organisms to avoid punishment (Pun), as alleles for punishers and non-punishers maintained similar frequency levels in the population.

The simulation given in column 2, row 2, is similar to the one described before, but organisms were kept shameless (Sha=0) and cost of punishment were paid only by the punished and never by the punisher. Here, costs of punishment did not evolve towards minimum values (Cost) and the balance between punishers and non-punishers was equilibrated, similar to the one of the simulation with shameful individuals. The levels of generosity (Gen) were high, avoiding punishment.

In colum2, row 1, a simulation where organisms could have shame and exercise punishment without a cost to the punisher, showed again that evolution maintained a balance between punishers and non-punishers (Pun), and that the levels of generosity (Gen) were high so as to avoid punishment.

In the simulation given in column 3, row 1, the cost of punishment was not free, as punishers had to pay the same cost at that received by the punished. Here, the levels of generosity (Gen) tended to be low and organisms tended to avoid being punishers (Pun=0).

When simulations allowed organisms to show shame (column 1, row 2), or to engage in punishment costless to the punisher (Column 2, row 2), punishers and non-punishers coexisted in harmonious equilibrium. But only in the last case did evolution favored pro-social behavior, i.e. a large proportion of individuals with alleles for non-zero generosity.

If organisms were allowed to acquire an allele that let them avoid engaging in non-commercial social interactions, most of them did so (ASt=0), as shown in the simulation given in column 3, row 2.

Row 3 shows simulations where organisms could acquire three types of alleles regulating their non-commercial social interactions. Allele 0 helped them to avoid these interactions, allele 1 made donations to be altruistic, and allele 2 converted the donation into a mutualistic act, as organism's acquired more "spices" when donating "sugar". In this situation, both altruistic punishment (Cost>0) and/or shame (Sha>0) drove evolution towards populations with high frequency of generous individuals (Gen>0), high proportion of punishers (Pun=1), and a non trivial amount of organisms engaging in non-commercial social behavior (ASt>0).

Repeated simulations (2000) showed that these results were statistically significant ($p<0.0001$)

Discussion

The simulations revealed that a gene predisposing for shame that emerges in evolution can easily invade a population, as it helps individuals to avoid social punishment. The adaptive advantage of shame is based on the fact that it increases flexibility to the shameful individual, allowing it to act selfish if the probabilities of being punished are low and achieving a reduction in the costs of social punishment if frequent punishment is likely. Cooperation, shame and punishment, as simulated here, produce a fluctuating dynamics of social cooperation, achieving long periods where the populations stabilizes pro-social behavior interspersed with periods where selfish behavior predominates. This temporal stabilization of pro-social behavior might provide societies with sufficient time to built institutions that might sustainable stabilize pro-social behavior. This dynamics resembles chaotic fluctuations between two attractors, where most individuals either cooperate or show selfish behavior. The results suggest that this complex dynamics should be more common in social systems than hitherto recognized. This has been proposed for the evolution of cooperation before [32] but applies to many other situations. For example, selfish punishers have been shown to self-limit their populations, causing a fluctuating population dynamics [33]. Even yeast show oscillations of morpho-types in population in social conflict situations [34].

The existence of shame and punishment alone, as explored here, can not explain the emergence of social behavior. Other approaches must be taken to understand the dynamics underlying the maintenance of social cooperation [11, 13, 15, 35, 36]. But once social cooperation exists, and if social punishment is applied, shame will emerge spontaneously.

Costless behaviors are not likely to occur in nature. Yet, if shame works pre-emotively to avoid social punishment before the individual receives punishment, costs involved in punishment can be kept to a minimum. In such a scenario, shame not only reduces costs, but will increase the fitness of the shameful agent by avoiding full payment of punishment costs. Thus, in real situations, shame may be much more likely to be maintained and to have a positive effect on the maintenance of cooperation than in the present simulations where conditions were purposefully chosen to not favor the shameful.

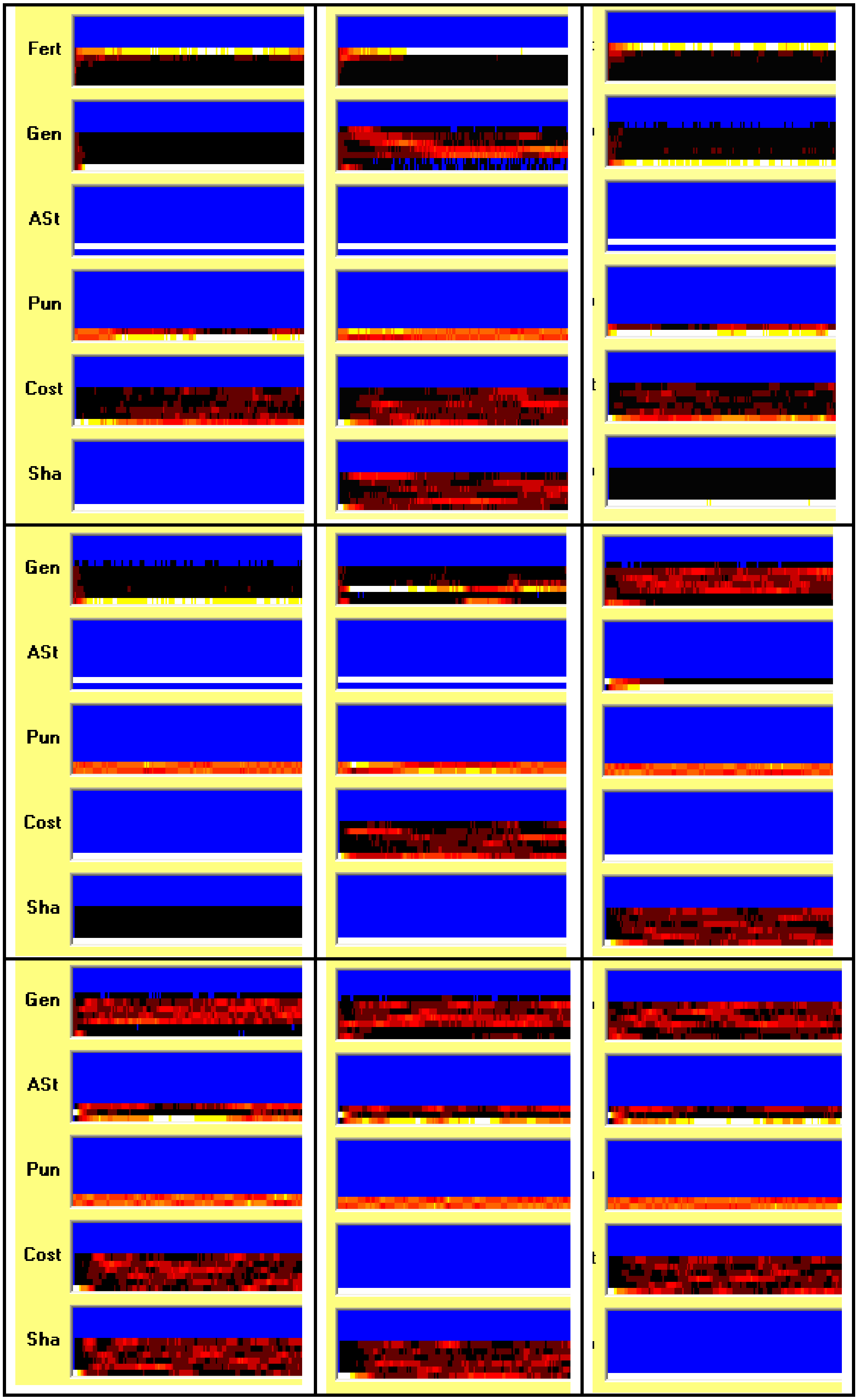
Shame, guilt and regret [26, 37] seem to be much more common among human societies, and may be among other social animals, than hitherto recognized. The findings that shame can easily evolve through individual selection without the need for any further assumptions, might have important implications for management of social environments. Shaming is shown to be a very effective strategy to favor socially desirable behaviors. It might work even better than costly punishment.

References

1. Axelrod, R.; Hamilton, W. The Evolution of Cooperation. Science 1981, 211: 1390–1396.
2. Bowles, S.; Gintis, H. The Origins of Human Cooperation. In The Genetic and Cultural Origins of Cooperation, Hammerstein, Peter, Ed., MIT Press, Cambridge, 2003.
3. Gintis, H. The Hitchhiker's Guide to Altruism: Gene-Culture Coevolution, and The Internalization of Norms. J. of Theoretical Biology 2003, 220: 407-418.
4. Johnson, D.D.P.; Stopka, P.; Knights, S. The Puzzle of Human Cooperation. Nature 2003, 421: 911–912.
5. Sachs, J.L.; Mueller, U.G.; Wilcox, T.P.; Bull, J.J. The Evolution of Cooperation. Q. Rev. Biolo. 2004, 79: 135-160.
6. Burtsev, M.; Turchin, P. Evolution of Cooperative Strategies From First Principles. Nature 2006, 440: 1041-1044.
7. Hauert, C.; Holmes, M.; Doebeli, M. Evolutionary Games and Population Dynamics: Maintenance of Cooperation in Public Goods Games. Proc Biol. Sci. 2006, 273: 3131–3132.
8. Williams, G. C. Adaptation and Natural Selection, (New Jersey: Princeton University Press, 1962 (Edition 1992).
9. Wilson, D.S. Altruism and Organisms: Disentangling the Themes of Multilevel Selection Theory. Am. Nat. 1996, 150 (suppl.): 122-134.
10. Maynard Smith, J. Group Selection and Kin Selection. Nature 1964, 201: 1145-1147.
11. West, S.A.; Griffin, A.S.; Gardner, A. Evolutionary Explanation for Cooperation. Current Biology 2007, 17: 661-672.
12. Hamilton, W. D. The Genetic Evolution of Social Behavior. J. of Theoretical Biology 1964, 7: 17-18.
13. Nowak, M.A.; Sigmund, K. Evolution of Indirect Reciprocity by Image Scoring. Nature 1998, 393: 573.
14. Jaffe, K. An Economic Analysis of Altruism: Who Benefits from Altruistic Acts? J. of Artificial Societies and Social Simulations 2002 5: 3.
15. Panchanathan, K.; Boyd, R. Indirect Reciprocity can Stabilize Cooperation without the Second-Order Free Rider Problem. Nature 2004, 432, 499-502.
16. Ohtsuki, H.; Hauert, C.; Lieberman, E.; Nowak, M.A. A Simple Rule for the Evolution of Cooperation on Graphs and Social Networks. Nature 2006, 441: 502-505.
17. Wilson, E.O. Sociobiology, The New Synthesis. Harvard Press, 1975.
18. Jaffe, K. On the Relative Importance of Haplo-Diploidy, Assortative Mating and Social Synergy on the Evolutionary Emergence of Social Behavior. Acta Biotheoretica 2001, 49: 29-42.
19. Boyd, R.; Richerson, P.J. Punishment Allows the Evolution of Cooperation (or Anything Else) in sizable groups. Ethology and Sociobiology 1992, 13: 171-195.
20. Sigmund, K.; Hauert, C.; Nowak, M. A. Reward and Punishment, Proceedings of the National Academy of Science 2001, 98: 10757-10762.
21. Fehr, E.; Gächter, S. Altruistic Punishment in Humans. Nature 2002, 415: 137-140.
22. Brandt, H.; Hauert, C.; Sigmund, K. Punishment and Reputation in Spatial Public Goods Games. Proceedings of the Royal Society Series B 2003, 270:1099-1104.
23. Brandt, H.; Hauert, C.; Sigmund, K. Punishing and Abstaining for Public Goods. Proceedings of the National Academy of Sciences of the United States of America 2006, 103:495-497.
24. Boyd, R.; Gintis, H.; Bowles, S.; Richerson, P.J. The Evolution of Altruistic Punishment. Proceedings of the National Academy of Science 2003, 100: 3531-3535.
25. Jaffe, K. Altruistic Punishment or Decentralized Social Investment? Acta Biotheoretica 2004, 52: 155-172.

26. Rawlings, E.I. Witnessing Harm to Other: A Reassessment of the Role of Guilt in Altruistic Behavior. Journal of Personality and Social Psychology 1968, 10: 377-80.
27. Jaffe, K. Simulations Show that Shame Drives Social Cohesion. IBERAMIA/SBIA 2006 in: LNAI 4140: 88-97, Springer Verlag, Berlin. J.S. Sichman et al. (Eds), 2006.
28. Jaffe, K. From Sociodynamics to Synergistics: Understanding the Wealth of Nations Using Simulations. Proceedings of ESSA 2004, Valladolid, Spain 2004.
29. Jaffe, K. Monte Carlo Exploration of Mechanisms for the Creation of Aggregate Wealth. Proceedings of IAREP/SABE 2002 Conference, Turku, Finland, 2002.
30. Jaffe, K. The Evolutionary Dynamics of Sex: Why the Sexes Are, In Fact, Always Two? Interciencia, 1996, 21: 259-267.
31. Jaffe, K. Emergence and Maintenance of Sex Among Diploid Organisms Aided by Assortative Mating. Acta Biotheoretica 2000, 48: 137-147.
32. Nowak, M.; Sigmund, K. Chaos and the Evolution of Cooperation. Proc. Natl. Acad. Sci. USA 1993, 90: 5091-5094.
33. Eldakar O.T.; Wilson D.S. Selfishness as Second-order Altruism. PNAS 2008, 105: 6982-6986,.
34. MacLean, R.C.; Gudelj, I. Resource Competition and Social Conflict in Experimental Populations of Yeast. Nature 2006, 441: 498-501.
35. Simon, H. A Mechanism for Social Selection and Successful Altruism, Science 1990, 250: 1665–1668.
36. Nowak, M. Five Rules for the Evolution of Cooperation. Science 2006, 314: 1560-1563.
37. Marchiori, D.; Warglien, M. Predicting Human Interactive Learning by Regret-Driven Neural Networks. Science 2008, 319: 1111-1113.

Figure 1: Nine different simulations of 500 time steps each. The figure gives the frequency of each allele for each gene in the population. The vertical axis indicates the type of allele from 0 to 10, as defined in Table 1. Frequency is expressed in colors: 0% of the total in the population (blue), less than 10% (black), over 50% (red), over 80% (yellow), and over 90% (white). The horizontal axis gives the results for different time steps of the simulation. Detailed explanations are given in the text.


Fert
Gen
ASt
Pun
Cost
Sha
Gen
ASt
Pun
Cost
Sha
Gen
ASt
Pun
Cost
Sha