When reputation matters, punishment may be reduced to the extreme cases

December 7 2006

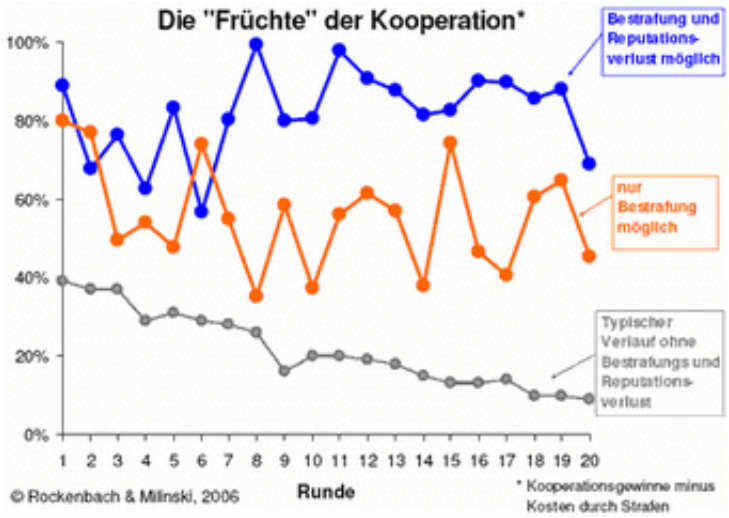

Fig. 1b: Profit through co-operation in the "Public Goods Game". Blue: if both punishment as well as loss of reputation are possible. Orange: if only punishment is possible. Grey: typical sequence without punishment and loss of reputation. Image: Rockenbach & Milinski

Scientists at the Max Planck Institute of Limnology and the University of Erfurt have discovered that even in the presence of effective reputation mechanisms, people do not entirely cease punishing those that put self-interest ahead of the collective interest. When given a choice, subjects even prefer a society with the option of both maintaining their own reputation and punishing "free riders". Acts of punishment are rare but imposed in a targeted manner.

Scientists therefore conclude that although punishment activities are as integral to human societies as our desire to maintain a reputation, sanctions can be limited to the "worst offenders" if reputation is also an important social value. (*Nature*, December 7, 2006).

Many current problems in human societies or states, such as the overexploitation of fish stock or the abuse of social welfare systems, represent a failure of co-operation. Such conflicts between social and individual interests over the use of resources, a phenomenon known as the "tragedy of the commons", could conceivably become a threat to the actual survival of humanity.

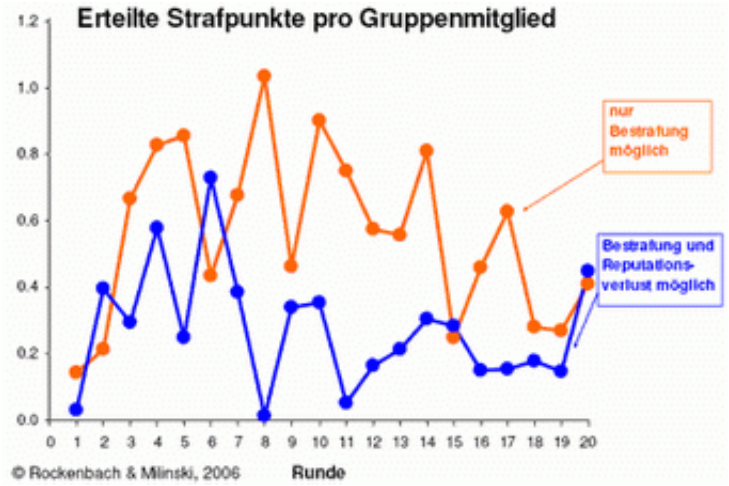

Fig. 1a: Average of penalty points per group member in each of the 20 rounds. Blue: if both punishment as well as loss of reputation are possible. Orange: if only punishment is possible. Image: Rockenbach & Milinski

One example of such a danger in our time is the unrestricted use of fossil fuels and its effects on the global climate. Scientific research has shown, however, that there are also factors that increase our willingness to co-operate such as directly punishing those exploiting a public resource for

personal gain or rewarding those with a high social status (reputation). The interesting question from a political perspective would be to what extent effective reputation mechanisms could render redundant costly sanctions of defectors.

To study such social dilemmas, experimental scientists use so-called "public goods games" as a testing ground. In a typical set up, four players are asked to contribute one Euro each to a common pool of funds. The experimenter then doubles the amount in the pool and distributes it equally among all participants - irrespective of whether they have actually made a contribution or not. If all players donated e.g. one Euro, they will end up with two Euros each. But with just one player defecting, the average dividend is reduced to 1.50 Euros each; a net profit of 1.50 Euros for the "free rider" and 0.50 Euro for the co-operator. At the end of the game, each participant will be paid his or her actual earnings in cash. While such experiments usually start out cooperatively, co-operation typically breaks down after only a few repetitions, with everyone ceasing to contribute to the public good, just as in real life. Any attempt to co-operate would entail a financial loss, since the investment of one Euro yields only a return of 50 Cents to the investor.

If, on the other hand, players are given the option of punishing non-cooperators - at incurring costs to themselves - the willingness of defectors to co-operate will increase within subsequent game rounds. In this scenario, however, financial gains are reduced, since both the punisher and the punished are losing money. As a result, the group would be better off if the willingness to co-operate was maintained by less harmful means.

In fact it is possible to sustain high levels of co-operation if participants have the opportunity to build up a good reputation and gain personal prestige through their behaviour. According to the motto "give and you shall receive", individuals that have supported others will receive support

in return. To achieve such "indirect reciprocity", it is important to build up a high donor status. At the same time, non-cooperative players are disciplined by denying them support, which acts as a punitive measure.

On the basis of these findings, one would expect that costly punishments would no longer be performed in social dilemma situations if there was an opportunity to build up a reputation by contributing to the common pool. After all, this is a significantly cheaper and more effective alternative to the sanctions. "We now know that punishing hardly pays off - the costs incurred both for the punisher and the punished will offset most of the gains resulting from enhanced co-operation", says Manfred Milinski from the Max Planck Institute of Limnology in Plön. The evolutionary biologist, together with economic scientist Bettina Rockenbach of the University of Erfurt, investigated what happened if players were not only given the options of costly sanctions but also of building a reputation. Would sanctions be replaced by reputation?

The scientists set up an experiment with students from the University of Erfurt. They divided participants into groups of eight people in order to play 20 rounds of a public good game with two different game options. Before the start of each round, players could either choose a group with an opportunity to both sanction fellow players and build up a reputation, or a group with reputation building opportunities only. In a second experiment, players were given the choice to either join a group with only the opportunity to sanction, or a group in which the game was played without either sanction or reputation building opportunities.

The findings were surprising: on the outset of both games, seventy percent of participants chose the option with no punishment opportunities. Although the participants in the first experiment had the opportunity of exclusively using reputation building in order to sustain co-operation levels - a solution at no expense and saving costs for the punishers - in subsequent rounds, more and more participants decided to

switch to the group that offered the combination of both punishment and reputation building. In this constellation, in comparison with the exclusive punishment variation of the game, players were punished less often (Fig. 1a), and the risk of being penalised steadily decreased. Punishment, however, remained a deterrent: in the rare instances where punishment was exercised, free riders were punished significantly more harshly than in the pure punishment control. "It is possible to maintain co-operation by exercising virtually no punishment if others are watching us and our own reputation is at stake", explains Bettina Rockenbach. As costly punishments are therefore reduced by two thirds, the public goods game also becomes enormously efficient: compared to the other game options, the contribution made to the common pool was the largest in the course of the game rounds (Fig. 1b).

Both punishing others and building a reputation for oneself seem to be key ingredients of human social behaviour. In addition, effective reputation mechanisms create an environment in which actual sanctions of social misconduct are progressively unnecessary. It stands to reason that such findings could be applied to some of our society's problems between people, groups, and states.

Citation: Bettina Rockenbach & Manfred Milinski, The efficient interaction of indirect reciprocity and costly punishment, *Nature*, December 7, 2006

Source: Max-Planck-Institute for Limnology