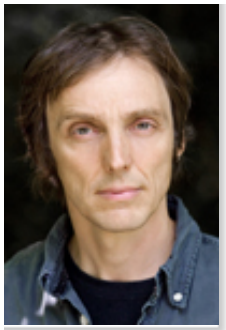
Published online 14 July 2010 | Nature | doi:10.1038/news.2010.351

Column: Muse

Who should pay for the police?



The punishment of antisocial behaviour seems necessary for a stable society. But how should it be policed, and how severe should it be? Game theory offers some answers, Philip Ball finds.

Philip Ball

A key axis of political opinion in democratic nations measures the 'size' of government. How much or how little should the state interfere in our lives? At one end of the axis sits political philosopher Thomas Hobbes, whose state is so authoritarian — absolute monarchy — that it barely qualifies as a democracy at all once the ruler is elected. At the other extreme is Peter Kropotkin, the Russian revolutionary anarchist who argued in his 1902 book *Mutual Aid* that people can organize themselves harmoniously without any government at all.

I'm prepared to guess that most *Nature* readers, being benign moderates, will cluster around the middle ground defined by John Stuart Mill, who argued that government is needed to maintain social stability, but should intrude only to the extent of preventing individuals from harming others.

The trouble is that 'harming others' is a slippery concept, illustrated most profoundly by the problem of the 'commons'. If you drop litter, if you don't pay your taxes or if you tip your sewage into the river, it's hard to pinpoint how or who your actions 'harm', if anyone. But if everybody does it, society suffers. So laws and penal codes must not only prevent or punish heinous crimes, but also discourage free-riders who cheat the mechanisms that promote social order.

Easy riders

Game theory and behavioural economics are now exploring how we collectively negotiate these rules. 'Public goods games', in which participants seek to maximize their rewards through competition or cooperation, have shown that people care about punishment to an irrational degree[1]. Say, for example, players are invited to put some of their money into a collective pot that will then be multiplied and divided among the players. The more the players put in, the better the pay-off. But if one person doesn't contribute, they still get the reward — so there's a temptation to free-ride.

Police officers — a natural consequence of evolution?

Julian Love/JAI/Corbis

If players are allowed to fine free-riders, but at a cost to themselves, they will generally do so: they care more about fairness than profit. This, however, can introduce another problem, a second-order temptation to free-ride: you contribute to the pot but leave others to shoulder the cost of sanctioning the cheaters who don't. So there's an infinite regress of opportunities to free-ride, which can eventually undermine cooperation.

But what if the players can share the cost of punishment by contributing to a pool in advance — equivalent, say, to paying for a police force and penal system? This decreases the overall profits — it costs society — because the pool goes to waste if no one cheats. Yet in a paper in *Nature* today[2], game theorist Karl Sigmund at the University of Vienna and his colleagues show in a computer model that pool-punishment can nevertheless evolve as the preferred option over peer-punishment as a way of policing the game and promoting cooperation: a preference, one might say, for a state police force as opposed to vigilante justice. This arrangement is, however, self-organized *à la* Kropotkin, not imposed from the top down *à la* Hobbes: pool-punishment simply emerges as the most stable strategy.

Crime and punishment

Of course, what often distinguishes these things in real life is that state-sponsored policing is more moderate and less arbitrary or emotion-led than vigilante retribution. That highlights another axis of political opinion: are extreme punishments more effective at suppressing defection than less severe ones? A modelling study of public goods games by Dirk Helbing at the Swiss Federal Institute of Technology Zurich (ETH Zurich) and his co-workers, soon to be published in the *New Journal of Physics*[3] and elaborated in another recent paper[4], suggests that fostering cooperation may depend on the strength of punishment in subtle, non-intuitive ways.

Above a critical punishment (fine) threshold, cooperators who punish can gain strength by sticking together, eventually crowding out both defectors and non-punishing cooperators (second-order free-riders). But if punishment is carried out not by cooperators but by other defectors, too high a fine is counterproductive and reduces cooperation. Cooperation can also arise through an unholy alliance of cooperators and defectors who both punish.

Why would defectors punish other defectors? This behaviour sounds bizarre, but is well documented experimentally[5], and familiar in reality: there are both hypocritical 'punishing' defectors (think of television evangelists whose condemnation of sexual

misdemeanours ignores their own) and 'sincere' ones, who deplore certain types of cheating while practising others.

In recent years, one of the most important lessons of these game-theory models has been that the outcomes are not necessarily permanent or absolute. What most people presumably want is a society in which people cooperate. But different strategies for promoting that have different vulnerabilities to an invasion of defectors. And strategies evolve: prolonged cooperation might erode a belief in the need for (costly) policing, opening the way for a defector takeover. Which is perhaps to say that public policy should be informed but not determined by computer models. As Stephen Jay Gould has said, "There are no shortcuts to moral insight."[6] ■

References

1. Fehr, E. & Gächter, S. *Am. Econ. Rev.* 90, 980-994 (2000). | **ISI** |
2. Sigmund, K., De Silva, H., Traulsen, A. & Hauert, C. *Nature* doi:**10.1038/nature09203** (2010).
3. Helbing, D., Szolnoki, A., Perc, M. & Szabó, G. preprint available at: **http://arxiv.org/abs/1007.0431** (2010).
4. Helbing, D., Szolnoki, A., Perc, M. & Szabó, G. *PLoS Comput. Biol.* **6**, e1000758 (2010).
5. Shinada, M., Yamagishi, T. & Ohmura, Y. *Evol. Hum. Behav.* **25**, 379-393 (2004). | **Article**
6. Gould, S. J. Natural History 106 (6), 12-21 (1997).