SOCIAL SCIENCE

Carrot or stick?

Rewards and punishments can cajole people into cooperating, but they are costly to implement. A theoretical study finds that, when participation in group activities is optional, punishing uncooperative behaviour is the cheaper method.

SIMON GÄCHTER

The philosopher John Locke[1] once wrote, "Good and evil, reward and punishment, are the only motives to a rational creature". Although Locke was referring primarily to the discipline of children, reward and punishment are motivational forces for behaviour across many domains of social life. Understanding the consequences of such 'carrots and sticks' is a core topic in the behavioural sciences, particularly in studies of cooperation[2–8]— behaviour that benefits others or the group at a cost to the cooperating individual. Many problems in modern human societies, from interactions in the workplace to tackling climate change, require genetically unrelated individuals to cooperate in situations in which collective welfare is jeopardized by individual self-interest. So how do rewards and punishments curb selfishness and help to maintain social order? A paper by Sasaki *et al.*[9] in *Proceedings of the National Academy of Sciences* helps to answer these questions.

Theoretical and experimental research on the evolution of cooperation has concentrated on punishment, with relatively few studies investigating reward[7,10–12]. Furthermore, most studies have focused on 'peer punishment', in which defectors are punished by group members[3,7,8,10–14]. However, the findings of peer-punishment studies may not be broadly applicable to modern human societies, which have developed formal sanctioning systems, whereby rewards and punishments are carried out by rule-bound institutions rather than by

individuals (Fig. 1). Sasaki *et al.* address this with a theoretical assessment of how institutionalized reward and punishment systems regulate cooperative behaviour, and at what relative cost.

The authors used evolutionary game theory for their analysis — a theoretical framework in which strategic behaviours can be analysed in the context of evolutionary selection pressure. Their analysis used the 'public goods game', in which people can either cooperate or defect, and in which cooperation is collectively beneficial, but defection is better for self-interest. People learn by observing others and they emulate successful individuals, so that strategies that yield greater pay-offs proliferate. The model in this study compares one institution that rewards cooperators and another that punishes defectors. The authors studied reward and punishment under two conditions: in one, all individuals were forced to take part; in the other, participation was voluntary.

C. SCHMIDT/CORBIS

Figure 1 | Punishment comes cheap. In modern societies, rule-bound institutions, such as the legal and education systems, use punishments and rewards to encourage cooperative behaviour. Sasaki and colleagues' theoretical analysis[9] shows that, when individuals can choose whether or not to participate, punishment is the less costly method.

Sasaki and colleagues found that during compulsory participation the two incentives (punishment or reward) lead to the same outcomes if they are very small or very large. It seems that if either of the incentives is too small, cooperation cannot be achieved because a population of cooperators can be invaded by defectors. If they are large enough, both types of incentive can lead to a population of cooperators. However, differences arise when the incentives are of only intermediate value. Punishments of intermediate severity produce stable populations of either defectors or cooperators, whereas rewards of intermediate value lead to stable mixed populations in which only partial cooperation is achieved.

The authors then changed the rules of the game to allow individuals to opt out, which causes the outcomes to change remarkably. The game now considers three strategies: non-participation, defection and cooperation. In this game, very small incentives lead to an unstable pattern of non-participation with bursts of cooperation. When the incentives are very large, a stable uniform population of cooperators emerges — as it did during forced participation. The most remarkable outcome of the study occurs when intermediate incentives are offered and participation is voluntary. In this situation, slightly increasing the severity of the punishment above a very low level results in stable populations of cooperators. By contrast, rewards of at least medium size are needed to cause a shift in the population from a majority who opt out of participation to a stable mixture of cooperators and defectors. When participation is voluntary, only very large rewards could generate a stable and uniform population of cooperators.

This noteworthy theoretical finding has practical implications. Sasaki and colleagues' model can be used to calculate the size of the incentive needed to achieve a stable level of cooperation, although such calculations will be difficult to apply in reality. However, the authors' analysis does raise several issues that are relevant for understanding modern human societies. First, they show that voluntary participation crucially influences the relative cost of reward and punishment. But in modern societies people can hardly opt out of the law. This might suggest that the results of games played with enforced participation are more applicable. Second, real-life institutions do not work perfectly, for example punishment or reward may not be correctly implemented. It remains unclear how these imperfectly applied incentives might affect cooperation. Third, law enforcement in reality typically relies on punishment rather than reward. This study may have identified a reason why punishment has become the default in societies — because punishment is a cheaper and more reliable way of inducing cooperation than is reward.

Other studies of human behaviour have also found that cooperation is strongly influenced by changes in the size of the incentive[13]. In experimental studies, reward and punishment induce similar levels of cooperation when the incentive is very large[11,12]. The threat of a strong punishment can achieve cooperation at a very low cost[14]. For an intermediate level of incentive, punishment can induce greater cooperation than reward[12], but not consistently so[10]. Finally, cooperation breaks down rapidly if both forms of incentive are removed[8].

Although these studies were conducted in settings of forced participation and peer punishment (or reward), their findings are encouragingly similar to Sasaki and colleagues' theoretical analysis[9]. These similarities suggest that experimental analyses conducted in a manner closer to the framework presented by Sasaki *et al.* — institutional delivery of incentives and assessment of voluntary participation — might further enhance our understanding of how reward and punishment maintain social order.

The political scientist Mancur Olson recognized the importance of reward and punishment for cooperation when he argued in his seminal 1965 study[15] that "the recalcitrant individual can be ostracized, and the cooperative individual can be invited into the center of the charmed circle." But Olson, Locke and other earlier scholars who wrote about reward and punishment in human social affairs relied on casual observation and introspection. Collectively, Sasaki and colleagues' study, the related experimental analyses and the potential investigations that arise from them are an example of how researchers today are much better equipped to combine rigorous theoretical and experimental analyses tounderstand sticks and carrots[16]. And this is a rewarding situation indeed. ■

Simon Gächter *is in the Centre for Decision Research and Experimental Economics, University of Nottingham, Nottingham NG7 2RD, UK.*
e-mail: simon.gaechter@nottingham.ac.uk

1. Locke, J. *Some Thoughts Concerning Education* (Cambridge Univ. Press, 1889).
2. Ostrom, E., Walker, J. M. & Gardner, R. *Am. Polit. Sci. Rev.* **86,** 404–417 (1992).
3. Fehr, E. & Gächter, S. *Nature* **415,** 137–140 (2002).
4. Sigmund, K. *Trends Ecol. Evol.* **22,** 593–600 (2007).
5. Boyd, R., Gintis, H. & Bowles, S. *Science* **328,** 617–620 (2010).
6. Bowles, S. & Gintis, H. *A Cooperative Species: Human Reciprocity and its Evolution* (Princeton Univ. Press, 2011).
7. Balliet, D., Mulder, L. B. & Van Lange, P. A. *Psychol. Bull.* **137,** 594–615 (2011).
8. Chaudhuri, A. *Exp. Econ.* **14,** 47–83 (2011).
9. Sasaki, T., Brännström, Å., Dieckmann, U. & Sigmund, K. *Proc. Natl Acad. Sci. USA* **109,** 1165–1169 (2012).
10. Sefton, M., Shupp, R. & Walker, J. M. *Econ. Inquiry* **45,** 671–690 (2007).
11. Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D. & Nowak, M. A. *Science* **325,** 1272–1275 (2009).
12. Sutter, M., Haigner, S. & Kocher, M. G. *Rev. Econ. Stud.* **77,** 1540–1566 (2010).
13. Nikiforakis, N. & Normann, H.-T. *Exp. Econ.* **11,** 358–369 (2008).
14. Gächter, S., Renner, E. & Sefton, M. *Science* **322,** 1510 (2008).
15. Olson, M. *The Logic of Collective Action* (Harvard Univ. Press, 1965).
16. Poteete, A. R., Janssen, M. A. & Ostrom, E. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice* (Princeton Univ. Press, 2010).