origin, rather than a fine-tuning to varied modern circumstances.

Second, the punishers in laboratory experiments such as Fehr and Gächter's are anonymous[5], so potential extra costs resulting from retaliation (of any sort) by victims are ruled out. Anonymity is unrealistic among early human groups: vigilantes would have to confront defectors to punish them, which incurs risk, and punishment among group members gives rise to grudges and reprisals, which undermine future cooperation. Although groups may be willing to punish individual defectors, people in one-on-one situations may not accept the personal cost of punishment (for example, they are often unwilling to intervene in criminal acts or to testify in trials for fear of retaliation).

Third, the problem remains of what prevents the occurrence of second-order free-riders, who cooperate for the public good but defect from bearing the cost of punishment[9]. Fehr and Gächter's results suggest that this is not a problem, as a core of people willingly incur personal costs to administer punishment, motivated by anger (although it is unclear whether they would act on it if they were not anonymous).

Alternative solutions are that punishment may come from an external institution, or it is not costly, or is administered to both defectors and individuals who fail to punish defectors. These alternatives have been criticized[9] but, if punishment is important, we suggest that an important source has been overlooked. The 'external' solution has been rejected because cooperation is prevalent in pre-industrial human groups, despite the absence of the enforcing institutions of modern states[9] — but this ignores religion, a feature of all human societies. Religions share taboos and codes of conduct that often promote cooperation for the public good and threaten supernatural punishment for those who do not follow these codes. Followers fear the personal consequences of defecting, and may be prepared to be altruistic if they believe that those who are not will be punished (now or in an afterlife).

A belief in supernatural punishment may arise as an abstract product of culture but thereafter become subject to selection (that is, the argument does not rely on invoking evolutionary origins for religious beliefs). But such beliefs could have evolved, either by group selection[8] or natural selection — as with the 'green beard' effect, individuals who signal common attributes can cooperate selectively with each other and thus outperform others. Groups with costly religious beliefs that signal commitment and loyalty outlive non-religious groups as a result of improved cooperation[10]. A 'stick' may be a good way to coerce people into cooperating, but the 'carrots' of kin cooperation, reciprocal altruism, reputation-building and religion have been crucial alternatives over our long evolutionary history, with a legacy that pervades today.

**Dominic D. P. Johnson*, Pavel Stopka†, Stephen Knights‡**

**Olin Institute for Strategic Studies, Harvard University, Cambridge, Massachusetts 02138, USA*
*e-mail: dominic@post.harvard.edu*
†*Department of Zoology, Charles University, 128 44 Prague 2, Czech Republic*
‡*Department of Economics, University of Oxford, Oxford OX1 3UQ, UK*

1. Henrich, J. *et al. Am. Econ. Rev.* **91,** 73–78 (2001).
2. Hamilton, W. D. *J. Theor. Biol.* **7,** 1–52 (1964).
3. Trivers, R. L. *Q. Rev. Biol.* **46,** 35–57 (1971).
4. Alexander, R. D. *The Biology of Moral Systems* (Hawthorne, Aldine, New York, 1987).
5. Fehr, E. & Gächter, S. *Nature* **415**, 137–140 (2002).
6. Clutton-Brock, T. H. & Parker, G. A. *Nature* **373,** 209–216 (1995).
7. Ostrom, E., Walker, J. & Gardner, R. *Am. Polit. Sci. Rev.* **86,** 404–417 (1992).
8. Sober, E. & Wilson, D. S. *Unto Others: The Evolution and Psychology of Unselfish Behaviour* (Harvard Univ. Press, Cambridge, Massachusetts, 1998).
9. Henrich, J. & Boyd, R. *J. Theor. Biol.* **208**, 79–89 (2001).
10. Sosis, R. *Cross-Cult. Res.* **34,** 71–88 (2000).

*Fehr and Gächter reply* — The claim by Johnson *et al.* that human cooperation in social-dilemma games violates rational-choice theory is not justified[1]. If people have altruistic aims, altruistic behaviour is a rational means by which to achieve their proximate goals. From an evolutionary viewpoint, we need to explain why humans are often altruistic by strong reciprocity[2–4]. Although kin selection, reciprocal altruism and indirect reciprocity explain relevant forms of human cooperation[5–7], they do not ultimately explain strong reciprocity[8].

Kin selection would account for strong reciprocity if human behaviour were driven by rules that do not distinguish between kin and non-kin. But humans, like other primates, distinguish cognitively and behaviourally between the two[5,9], and generally feel stronger emotions towards kin. Likewise, reciprocal altruism could account for strong reciprocity if humans' behavioural rules did not depend on the probability of future interactions with potential opponents. But humans can distinguish long-term partners from people with whom future interaction will be less likely ('strangers'), and will cooperate more if they anticipate that interaction will be frequent[6]. Emotional responses may also be stronger towards a long-term partner than towards a 'stranger' (our unpublished results).

Reputation-based ultimate theories could account for strong reciprocity if our behavioural rules did not depend on our actions being observed by others. However, if reputation formation is ruled out, cooperation breaks down, whereas it flourishes if subjects gain in reputation[7].

Early humans whose behaviour was fine-tuned to respond to kin or non-kin, partners or strangers, and gaining in reputation, probably had an evolutionary advantage because, contrary to common belief, they faced interactions where the probability of future encounters was sufficiently low as to make defection worthwhile. Ethnographic evidence indicates that humans had many encounters with individuals with whom they had little future interaction[8]. In addition, the costs of mistakenly treating unrelated individuals as kin, or treating strangers as partners, were high — for instance, a lack of vigilance with strangers could be fatal. Because of these costs, individuals who could adjust their behaviour to suit the their opponent's characteristics had greater fitness. The problem with any theory claiming that strong reciprocity is maladaptive in modern circumstances is that individuals understand the risks of exploitation in interactions with non-kin and strangers, and behave accordingly. An evolutionary explanation of strong reciprocity is needed that does not assume that individuals are maladapted[2,3].

A proximate mechanism of belief in supernatural punishment does not solve the evolutionary puzzle. How could such beliefs evolve if those who did not hold them defected and hence gained an advantage? Laboratory experiments do not support the claim that religion is important for cooperation. If other people in the group are expected to defect, then almost everyone else — religious or not — will defect too[10]. Moreover, in almost all religions, non-believers have been ostracized and have faced worldly punishment.

We do not agree that anonymity is a problem in the experiment: it rules out other, less costly forms of social punishment that are available in non-anonymous situations, such as workers' hostility towards strike-breakers and people's hostility towards wartime deserters. If non-anonymous punishment were lessened by being more costly, this could be just another example of how remarkable humans are at fine-tuning their behaviour to suit their circumstances.

**Ernst Fehr*, Simon Gächter†**

**Institute for Empirical Research in Economics, University of Zürich, 8006 Zürich, Switzerland*
*e-mail: efehr@iew.unizh.ch*
†*University of St Gallen, FEW-HSG, 9000 St Gallen, Switzerland*

1. Gintis, H. *Game Theory Evolving* (Princeton Univ. Press, 2000).
2. Henrich, J. & Boyd, R. *J. Theor. Biol.* **208,** 79–89 (2001).
3. Gintis, H. *J. Theor. Biol.* **206,** 169–179 (2000).
4. Fehr, E. & Gächter, S. *Nature* **415,** 137–140 (2002).
5. Silk, J. B. *Am. Anthropol.* **82,** 799–820 (1980).
6. Gächter, S. & Falk, A. *Scand. J. Econ.* **104,** 1–25 (2002).
7. Milinski, M., Semmann, D. & Krambeck, H. J. *Nature* **415,** 424–426 (2002).
8. Fehr, E. & Henrich, J. in *The Genetic and Cultural Evolution of Cooperation* (ed. Hammerstein, P.) (MIT Press, Cambridge, Massachusetts, in the press).
9. Tomasello, M. & Call, J. *Primate Cognition* (Oxford Univ. Press, New York, 1997).
10. Fischbacher, U., Gächter, S. & Fehr, E. *Econ. Lett.* **71,** 397–404 (2001).