

Colman, A. M. (2006). The puzzle of cooperation: Review of H. Gintis, S. Bowles, R. Boyd, & E. Fehr, *Moral sentiments and material interests: The foundations of cooperation in economic life*. *Nature*, 440, 744-745.

The puzzle of cooperation

Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life

edited by Herbert Gintis, Samuel Bowles, Robert Boyd & Ernst Fehr
MIT Press: 2005. 404 pp. \$50.£32.95

Andrew M. Colman

Robert May began his last presidential address to the Royal Society on 30 November 2005 by saying: “The most important unanswered question in evolutionary biology, and more generally in the social sciences, is how cooperative behaviour evolved and can be maintained in human or other animal groups and societies”. For example, birds often emit alarm calls when they spot predators, but how could such behaviour have evolved? A mutant bird that never gave alarm calls would save energy and avoid additional risk to itself while enjoying the benefits of its conspecifics’ alarm calls. Its “selfish gene” should therefore spread to fixation in the population.

For the same reason, cooperation is difficult to maintain when individuals are tempted to defect. A recent human example in Britain is the decline in voluntary take-up of the combined measles–mumps–rubella (MMR) vaccination by parents wishing to avoid an alleged health risk to their own children while implicitly relying on enough other children getting vaccinated to maintain “herd immunity”. This has the strategic structure of a social dilemma, because if all parents followed this individualistic reasoning, then everyone could end up worse off than if all behaved cooperatively.

A similar social dilemma is devastating UK fish stocks: over-fishing destroyed British herring fisheries long ago and is now causing terminal decline in other fish stocks in the English Channel, the North Sea and the Baltic. Anyone who makes a living by fishing is motivated to catch as many fish as possible, because restraint is pointless if enough others are exercising restraint, and is futile if they are not. But then fish are driven to extinction and everyone is worse off than if they had all restrained themselves cooperatively.

Bill Hamilton’s theory of “inclusive fitness”, or kin selection, explains the evolution of cooperation among genetically related individuals. It can explain the extreme self-sacrificing cooperation of female social Hymenoptera, who have 75 per cent of their genes in common, but not cooperation among non-relatives. Trivers’ theory of reciprocal altruism shows how cooperation between non-relatives can evolve if its cost is small and is outweighed by favours returned in the future. These two theories go some way towards explaining how cooperation evolved, but neither can explain human cooperation in unrepeatable interactions between strangers.

To fill this gap, Ernst Fehr and Simon Gächter introduced in 2000 a version of the theory of strong reciprocity incorporating the “altruistic punishment” of non-cooperators. *Moral Sentiments and Material Interests* is devoted to their theory’s biological, anthropological, economic, and social ramifications and related ideas. An introductory chapter is followed by three on the behavioural ecology of cooperation, four on modelling and testing strong reciprocity, and five on reciprocity and social policy, all by researchers in the vanguards of their fields.

According to the theory, cooperation is necessary for the provision of public goods, and punishment of non-cooperators, or free-riders, is itself a public good – a service provided for the benefit of the whole community. Such punishment is altruistic because it is costly to those who administer it, as it takes time and energy and invites retaliation. Fehr and Gächter have provided persuasive experimental evidence, reviewed in the book, that cooperation flourishes when punishment is possible and breaks down rapidly when it is not.

A problem not addressed in the book is that, because altruistic punishment is costly, natural selection should tend to eliminate it. Failure to punish defectors must presumably be treated as second-order defection, itself subject to sanctions from other group members. But what about sanctions against third-order defectors who neglect to punish second-order defectors, and so on? This is an infinite regress that becomes less credible with the addition of each successive layer of explanation. Strong reciprocity is an important and illuminating discovery, but we seem to have replaced the problem of explaining cooperation with that of explaining altruistic punishment.

The book provides a superb interdisciplinary synthesis of cooperation as explained by strong reciprocity and associated phenomena. Other explanations of cooperation in unrepeated interactions between strangers hardly get a look in, however. The most important is Richard Alexander's theory of "indirect reciprocity", according to which people use observations of direct reciprocity between others when deciding how to act towards them in the future. People benefit by cooperating, even in one-off encounters with strangers, because cooperation enhances one's reputation for cooperativeness and elicits reciprocal cooperation from others. This is a powerful theory, supported by evidence from computational and experimental studies, but *Moral Sentiments and Material Interests* mentions it only in passing.

Andrew M. Colman is in the School of Psychology, University of Leicester, Leicester LE1 7RH, UK.