

rest of the band (non-kin and distant kin) can be seen as free-riders who benefit substantially but pay no costs.

However, there is another way setting up the genetic cost/benefit analysis. The executioner who pays such costs is merely caught in a structural position, in which he becomes the chosen executioner because he is close kin, whereas the free-rider roles of those who abstain are also determined by social position. Thus, free-rider *genes* are not at issue because the free-riding is determined *situationally*.

In this light, we may reconsider the ethnographically well-described Mbuti case Guala cites from Turnbull (1961). Cephu cheats on a meat-acquisition system which is designed to bring in a fair share of game for all the participating families; and, collectively, most of the band actively shames him in ways that are humiliating while Cephu's loyal followers stand aside – but do not actively back him. This too is situational, because they are kin. It is worth noting that the sanctioning goes beyond shaming when one band member threatens the arrogant Cephu with ejection from the band; but he is taking little risk because the people backing him are in a state of moral outrage.

I emphasize that the several families associated closely with Cephu likely would be conventionally modeled as free-riding defectors because they stand aside; and also, that in fact this is not a matter of opportunistic free-rider genes in action. It is simply a situational matter, and over the millennia such stepping aside has had nothing to do with genes. In such contexts, the free-rider problem does not apply.

Guala has opened up some interesting questions, and has used ethnographic data in doing so. Perhaps these further ethnographic nuances may serve as useful food for thought, for scholars who use experiments with students (or non-LPA nonliterates) to try to understand human nature.

#### ACKNOWLEDGMENT

I thank Joe Henrich for comments on this commentary.

## The punishment that sustains cooperation is often coordinated and costly

doi:10.1017/S0140525X1100118X

Samuel Bowles,<sup>a</sup> Robert Boyd,<sup>b</sup> Sarah Mathew,<sup>b</sup> and Peter J. Richerson<sup>c</sup>

<sup>a</sup>*Santa Fe Institute, Santa Fe, NM 87501;* <sup>b</sup>*Department of Anthropology, University of California—Los Angeles, Los Angeles, CA 90095;* <sup>c</sup>*Department of Environmental Science and Policy, University of California—Davis, Davis CA 95616.*

samuel.bowles@gmail.com    rboyd@anthro.ucla.edu  
smathew@ucla.edu    pjricherson@ucdavis.edu  
<http://www.santafe.edu/~bowles>  
<http://www.sscnet.ucla.edu/anthro/faculty/boyd/>  
<http://smathew.bol.ucla.edu/Site/Home.html>  
<http://www.des.ucdavis.edu/faculty/Richerson/Richerson.htm>

**Abstract:** Experiments are not models of cooperation; instead, they demonstrate the presence of the ethical and other-regarding predispositions that often motivate cooperation and the punishment of free-riders. Experimental behavior predicts subjects' cooperation in the field. Ethnographic studies in small-scale societies without formal coercive institutions demonstrate that disciplining defectors is both essential to cooperation and often costly to the punisher.

We are grateful to Francesco Guala for providing a thoughtful reflection on what recent social dilemma experiments can tell us about real-world cooperation and the need for complementary ethnographic, historical approaches. But Guala's contribution is packaged along with what we think is a misunderstanding of our work, an overly pessimistic appraisal of the external validity

of experimental results, and a very partial reading of the evidence on costly punishment in small-scale societies.

The core of strong reciprocity is that human cooperation cannot be understood entirely as the result of repeated social interaction and self-interested individual calculation. Instead, people are motivated to cooperate with one another and to punish free-riding by a variety of ethical and other-regarding motives. Guala gets this right. However, he incorrectly believes that strong reciprocity requires punishment to be both very costly and uncoordinated. Punishment is costly when the cost of administering punishment, however small, exceeds the private benefit it creates for the punisher, thus giving rise to a second-order free-rider problem. Mechanisms like conformism, kin selection, or cultural group selection can solve the second-order free-rider problem, but usually only if the cost of punishment is low, either because it is rare (e.g., Boyd et al. 2003; Henrich & Boyd 2001) or because it is collectively administered (Boyd et al. 2010).

Everyday social life, even among strangers, is regulated by many individual acts of uncoordinated punishment. We are all aware of the pain we experience when we are frowned upon in public places among strangers. However, we agree with Guala that more costly forms of punishment in natural settings are usually collective. We capture this in our paper “Coordinated Punishment of Defectors Sustains Cooperation...” (Boyd et al. 2010), which Guala cites but seems to have misunderstood. In this model, potential punishers signal their willingness to punish, but they punish free-riders only when enough fellow punishers have signaled. When there is no assortment, there are two possible evolutionary equilibria: a population without punishment or cooperation, and a population with a mix of punishers and non-punishers in which most actors cooperate. Mean fitness is higher when punishers are present. When we allow an empirically realistic degree of assortment in the population, punishment may proliferate even when rare; and when it does, it is altruistic.

We developed this model because we share Guala's dissatisfaction with the typical representation of punishment as an individual act rather than something deliberated on by groups and undertaken jointly (but see Ertan et al. 2009). Nonetheless, experiments make a major contribution by showing that the predispositions that motivate punishment are common in many populations. We agree with Guala that we need better tests of the external validity of these experimental results. But two kinds of evidence are encouraging.

First, behavior in experiments predicts subjects' cooperation in the field. Brazilian shrimpers use large plastic bucket-like contraptions in which holes are cut to allow the immature shrimp to escape, thereby preserving the stock for future catches. Because they can cut holes of any size, the fishermen face a real-world social dilemma. Large holes represent cooperation with other fishers; small trap holes are a form of defection, and – just as in the Public Goods Game – having small holes is the dominant strategy for a self-interested shrimper. Not surprisingly, those who contributed most in a public goods experiment were also those who cut larger holes in their traps (Fehr & Leibbrandt 2011). The effects, controlling for a number of other possible influences on hole size, are substantial.

Additional evidence of external validity comes from a set of experiments and field studies with 49 groups of herders of the Bale Oromo people in Ethiopia, who were engaged in forest commons management (Rustagi et al. 2010, which Guala cites). The most common behavioral type in the experiments, constituting a bit more than a third of the subjects, were “conditional cooperators” who responded positively to higher contributions by others. Controlling for a large number of other influences on the success of the forest projects, the authors found that groups with more conditional cooperators planted more trees. (See Bowles & Gintis [2011] for more evidence on external validity.)

Second, there is much evidence for costly third-party punishment among societies without formal coercive institutions. Mathew and Boyd (2011) present extensive quantitative data showing that punishment of cowardice and other forms of free-riding plays an important role in warfare among the Turkana, an acephalous African pastoral group. Community members decide whether a violation occurred, and if it has, corporal punishment is administered by the violator's age-mates, not those damaged by the violation. Punishing takes time and effort and may damage valuable social relationships.

Contrary to Guala, punishment has been observed in the simplest foraging societies. Among the Walbiri of Australia, for example, offenses like homicide, physical assault, sacrilege, adultery, and theft were punished by death, wounding with a spear or knife, or attack with a club or boomerang (Meggitt 1962, pp. 256–59). The local community determined whether the act was an offense, decided on the punishment, nominated the person to carry out the punishment, and appointed the people responsible for ensuring that the punisher does not face retaliation (p. 255).

In some cases, meting out punishment is very costly. Among Aranda foragers of the Central Desert in Australia, wrongdoers were sometimes executed. The elders collectively decided on the fate of the wrongdoer, and assigned a group of young men to carry out the execution. Strehlow (1970, pp. 117–18) describes two cases in which the violator's relatives did not think the execution was justified, and killed the young men who had carried it out. According to Strehlow, capital punishment of this nature occurred in all Central Australian tribes before colonial administration made them a criminal offense.

## Weak reciprocity alone cannot explain peer punishment

doi:10.1017/S0140525X11001191

Marco Casari

*Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy.*

marco.casari@unibo.it <http://www2.dse.unibo.it/casari/>

**Abstract:** The claims about (1) the lack of empirical support for a model of strong reciprocation and (2) the irrelevant empirical role of costly punishment to support cooperation in the field need qualifications. The interpretation of field evidence is not straightforward, and other-regarding preferences are also likely to play a role in the field.

Guala should be praised for having raised this debate about punishment experiments. I will focus on two main points. First, the target article claims that the empirical evidence on peer punishment is not enough to support theories based on strong reciprocity. As I argue below, behavior in peer punishment experiments cannot be entirely rationalized with self-regarding or weak reciprocity attitudes, and strong reciprocity is one model of other-regarding behavior among others currently under debate.

There is no lack of anecdotes about peer pressure and punishment in field settings, ranging from high school students to miners on strike (Francis 1985) to fishermen communities (Bromley 1992) to workplaces (Kandel & Lazear 1992). In the region studied in Casari (2007), costly punishment is still practised today. Recently, 1,800 young grapevines have been cut with pruning hooks and shears. Apparently two people acted overnight, causing damage in thousands of euros. In the last five years, there have been seven similar episodes in the same community. Generally the culprits remain unknown (Nardon 2011). The issue of peer punishment was raised after field research and was not born as a laboratory anomaly. Experiments helped to clarify the extent and drivers of peer punishment,

because field evidence is often hard to interpret. There are nuisance factors and measurement limitations: The interaction may be repeated, the fine-to-fee ratio unknown, or institutions to promote cooperation may be present. Controlled experiments are useful because they remove many of these limitations. One robust finding is the willingness of many people to pay a personal cost to inflict a punishment on others, especially on free-riders. This result persists in one-shot situations when the punisher incurs a material loss. As in other experiments, the data point toward the existence of a mix of motivations in economic decision making. While most subjects exhibit exclusively self-regarding motivations, there are others who also exhibit an array of other-regarding motivations.

Weak reciprocity is simply not enough to rationalize the existing experimental results on peer punishment. For instance, subjects do not treat peer punishment as a second-order public good, that is, they do not employ punishment mainly to provide incentives for the free-rider to contribute, as a weak reciprocity argument would suggest (Casari & Luini 2006; 2009). One can also experiment settings with indefinite repetition, where weak reciprocators can support cooperative outcomes through a rational strategy different than costly peer punishment. When four subjects indefinitely played prisoner's dilemmas in random pairs, more than half of the time cooperators targeted defectors with peer punishment (Camera & Casari 2009). Rational, self-regarding subjects had the alternative to support full cooperation through a simple grim trigger strategy. Instead, many still employed peer punishment. To sum up, experiments on peer punishment have shed light over important aspects of cooperative behavior that are likely to apply also in field situations. Yet, the existing evidence still leaves some deep questions open about the genetic versus cultural origin of other-regarding motivations; about the degree of external validity of experiments; and, about what model can fit the observed patterns of punishment with reasonable precision.

Guala's second main point is that cooperation in the field does not rely primarily on the forces uncovered in punishment experiments but is promoted by institutions that reduce the costs of decentralized punishment and facilitate the functioning of weak reciprocity mechanisms. I agree, although I will discuss two of Guala's related statements, which are based on unconvincing interpretations of the anthropological evidence: (1) Peer punishment does not occur in the field; (2) hence, it is irrelevant in a field setting. Guala argues that peer punishment is rarely employed and that some punishment acts are not costly, given that the cost to inflict punishment is claimed to be "low." In the literature, what matters is the fine-to-fee ratio of a punishment act, not simply the absolute cost of a punishment request. Moreover, sanctions ought not to be always large but, rather, graduated (Ostrom 1990). In the lab one observes a proportion between crime and punishment, that is, actions of full free-riding attract more punishment than actions of partial free-riding, and something similar may be expected in the field.

When extrapolating to field situations, one has to keep in mind that in laboratory experiments, people are forced to interact with others, have little control over the information flow, and have only few options available. In the field, people have multiple ways to inflict punishment and have strategies alternative to peer punishment. Instead of physically confronting a norm violator, a cooperator may decide to act to lower the cost to punish, to create institutions, or to move camp elsewhere. Hence, people can optimize over the many strategies available. A lower-than-expected frequency of peer punishment actions may simply reveal that there are better strategies in that situation, not that they are unavailable or irrelevant. For instance, speaking up against someone is costly because it exposes one to the risk of retaliation (Wiessner 2005), as whistle-blowers know. To avoid counter-punishment, in the field people may increase the level of anonymity by spreading gossip instead of reproaching