

The restorative logic of punishment: Another argument in favor of weak selection

Nicolas Baumard

Abstract: Strong reciprocity theorists claim that punishment has evolved to promote the good of the group and to deter cheating. By contrast, weak reciprocity suggests that punishment aims to restore justice (i.e., reciprocity) between the criminal and his victim. Experimental evidences as well as field observations suggest that humans punish criminals to restore fairness rather than to support group cooperation.

As Guala rightly notes, there is very little evidence that punishment plays a role in the stabilization of cooperation in small-scale societies. On the other hand, as he also notes, it is difficult to totally rule out the strong view of punishment as it is complicated to precisely assess the costs of punishment in the field (Are there really no costs in punishing others? Aren't there many hidden benefits for the individual who punish? etc.). There is, however, another way to disentangle the two views of punishment, namely, the forms that punishments take. Indeed, the two theories – the weak and the strong – make different predictions regarding the logic of punishment.

Group selection theory holds that punishment aims to promote the good of the group by sustaining cooperation and preventing cheating (Boyd et al. 2003; Fehr & Gächter 2002; Henrich & Boyd 2001). This implies that punishment should be calibrated to deter crimes and render them non-advantageous. Here, group selection

parallels the utilitarian doctrine of punishment, which contends that punishment should be used to deter crimes and maximize the good of society (Polinsky & Shavell 2000; Posner 1983). The utilitarian theory of punishment holds, for instance, that the detection rate of a given crime and the publicity associated with a given conviction are relevant factors in assigning punishments. If a crime is difficult to detect, the punishment for that crime ought to be made more severe in order to counterbalance the temptation created by the low risk of getting caught. Likewise, if a conviction is likely to get a lot of publicity, a law enforcement system interested in deterrence should take advantage of this circumstance by “making an example” of the convict with a particularly severe punishment, thus getting a maximum of deterrence for its punishment.

By contrast, individual selection predicts a “restorative” or “retributive” logic for punishment (Baumard 2011). Restorative logic holds that punishment aims to restore justice between the criminal and the victim – either by harming the criminal or by compensating the victim. In intuitive terms, people are punished because they “deserve” to be punished, and not because punishing them would be useful for the society at large.

This restorative logic is a direct consequence of the way cooperation has evolved among humans (Baumard 2010; Trivers 1971). Indeed, human beings belong to a highly cooperative species and get most of their resources from collective actions, solidarity, exchanges, and so forth. (Gurven 2004b; Hill & Kaplan 1999). In the ancestral environment, individuals were in competition to be recruited for the most fruitful ventures, and it was vital to share the benefits of cooperation in a mutually advantageous manner. If individuals took a bigger share of the benefits, their partners would leave them for more interesting partners. If they took a smaller share, they

would be exploited by their partners who would receive more than what they had contributed to produce. This competition to attract cooperative partners is thus likely to have led to selection for a “sense of fairness,” a cognitive device that motivates individuals to share the costs and benefits of social interaction in an impartial way (André & Baumard 2011). If cooperation is based on fairness, then crimes create an unfair relationship between the criminal and her victim, and people have the intuition that the criminal ought to compensate the victim or to be punished in order to restore justice.

It is worth mentioning that this theory does not mean that punishment should be absent in human societies. As Guala notes, modern societies have found many institutional ways to reduce the costs of punishments. Although these institutions are absent in smaller societies, justice can still be restored by individuals seeking to retaliate. Retaliation is indeed advantageous from an individual perspective and can indeed be found in many non-human species (Clutton-Brock & Parker 1995). As Evans-Pritchard noted, in societies where there is no penal system, “self-help, with some backing of public opinion, is the main sanction” (Evans-Pritchard 1940).

In this kind of situation, selfish and moral motives converge: Victims (or their allies) attack criminals to signal their strength and gain a reputation as persons who cannot be attacked without risk, and by doing so, they also punish the wrongdoers by allowing justice to be done. In line with this idea, people in small-scale societies distinguish between legitimate (and proportionate) retaliation and illegitimate (and disproportionate) retaliation (Fürer-Hameendorf 1967; Miller 1990). Retaliation is thus clearly limited by moral concerns: Within the group, it has to be proportionate to the prejudice. As the *Lex Talionis* says, “an eye for an eye, a tooth for a tooth,” but no more.

Individual selection thus clearly predicts some kind of punishment, and more importantly, it predicts that punishments should aim to a specific goal (restoring fairness) that differs from the utilitarian goal predicted by group selection (preventing wrongdoing). Experimental studies, relying on a variety of methodologies, suggest that punishments fit more individual selection than group selection. Indeed, when people punish harmdoers, they generally respond to factors relevant to a retributive theory of punishment (magnitude of harm, moral intentions) and ignore factors relevant to the group selection theory (likelihood of detection, publicity, or likelihood of repeat offending) (Baron et al. 1993; Baron & Ritov 2008; Carlsmith et al. 2002; Darley et al. 2000; Glaeser & Sacerdote 2000; Sunstein et al. 2000).

In line with these results, field observations have extensively demonstrated that, in keeping with this prediction, the level of compensation in stateless societies is directly proportional to the prejudice inflicted to the victim: For example, the wrongdoer owes more to the victim if the wrongdoer has killed a family member or eloped with a wife than if he has stolen animals or destroyed crops (Hoebel 1954; Howell 1954; Malinowski 1926). To conclude, punishment does not seem to be a group adaptation. It follows the logic of fairness rather than the interests of the group.

References

- André, J. B. & Baumard, N. (2011) The evolution of fairness in a biological market. *Evolution*. [NB]
- Baron, J., Gowda, R. & Kunreuther, H. (1993) Attitudes toward managing hazardous waste: What should be cleaned up and who should pay for it? *Risk Analysis* 13(2):183–92. [NB]
- Baron, J. & Ritov, I. (2008) The role of probability of detection in judgments of punishment. Unpublished manuscript. [NB]
- Baumard, N. (2010) *Comment nous sommes devenus moraux: Une histoire naturelle du bien et du mal*. Odile Jacob. [NB]
- Baumard, N. (2011) Punishment is a not group adaptation: Humans punish to restore fairness rather than to help the group. *Mind and Society* 10(1). [NB]
- Boyd, R., Gintis, H., Bowles, S. & Richerson, P. (2003) The evolution of altruistic punishment. *Proceedings of the National Academy of Science USA*. 100(6):3531–35. [NB]
- Carlsmith, K., Darley, J. & Robinson, P. (2002) Why do we punish? deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology* 83(2):284–99. [NB]
- Clutton-Brock, T. & Parker, G. (1995) Punishment in animal societies. *Nature* 373(6511):209–16. [NB]
- Darley, J., Carlsmith, K. & Robinson, P. (2000) Incapacitation and just deserts as motives for punishment. *Law and Human Behavior* 24(6):659–83. [NB]
- Evans-Pritchard, E. E. (1940) *The Nuer, a description of the modes of livelihood and political institutions of a Nilotic people*. Clarendon Press. [NB]

Fehr, E. & Gächter, S. (2002) Altruistic punishment in humans. *Nature* 415:137–40.

[NB]

Glaeser, E. L. & Sacerdote, B. (2000) *The determinants of punishment: Deterrence, incapacitation and vengeance*. Harvard Institute of Economic Research Paper No. 1894. <http://ssrn.com/paper=236443> [NB]

Henrich, J. & Boyd, R. (2001) Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology* (208):79–89. [NB]

Hoebel, E. A. (1954) *The law of primitive man: A study in comparative legal dynamics*. Harvard University Press. [NB]

Howell, P. (1954) *A Manual of Nuer Law: Being an account of customary law, its evolution and development in the courts established by the Sudan government*. Published for the International African Institute by the Oxford University Press. [NB]

Malinowski, B. (1926) *Crime and custom in savage society*. Rowman & Littlefield. [NB]

Polinsky, A. M. & Shavell, S. (2000) The economic theory of public enforcement of law. *Journal of Economic Literature* 38(1):45–76. [NB]

Posner, R. (1983) *The economics of justice*. Harvard University Press. [NB]

Sunstein, C., Schkade, D. & Kahneman, D. (2000) Do people want optimal deterrence? *Journal of Legal Studies* 29(1):237–53. [NB]

Trivers, R. (1971) Evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57. [NB]