How the second-order free rider problem is solved in a small-scale society

By Sarah Mathew*

Moralistic punishment enables human cooperation, but an outstanding question is why do people voluntarily sanction when they can obtain the benefits of punishment without being enforcers themselves. To address how decentralized societies solve this second-order free rider issue I examine why people punish among the Turkana, a population in Kenya in which informal peer sanctioning sustains participation in high-stakes interethnic warfare. Using vignette experiments I show that Turkana subjects express punitive sentiments towards second-order free riders and those who sanction irresponsibly. The prevalence of such meta norms regulating punishment reveal a possible pathway by which moralistic punishment could have evolved.

Voluntary sanctioning of free riders can sustain cooperation but an outstanding problem is to explain why people punish (Yamagishi, 1986). Punishment is costly to implement and so second-order free riders who do not sanction wrongdoers can obtain the benefits of cooperation without paying the cost of enforcement. I address this question by examining the norms regulating informal punishment among the Turkana, a decentralized pastoral population in Kenya who utilize informal peer sanctions to maintain high-stakes participation in interethnic warfare (Mathew and Boyd, 2011). Using vignette experiments I show that subjects express punitive sentiments towards second-order free riders, towards those who sanction excessively or unjustly, and towards those who retaliate against legitimate punishment. Punishment is perceived as legitimate if it is decided upon by collective consensus, and administered by individuals from a preordained group. The prevalence of such meta-norms regulating punishment illustrate how a natural peer sanctioning system works, and highlight a potential pathway for norm enforcement to evolve.

Sanctioning of second-order free riders can resolve the second-order dilemma in theory (Kameda, Takezawa and Hastie, 2003), but empirical evidence for such second-order punishment has been mixed. Some work suggests that punishers gain reputation benefits (Barclay, 2006), but when second-order punishment is allowed in experiments, many subjects often punish enforcers instead of second-order free riders (Cinyabuguma, Page and Putterman, 2006; Herrmann, Thoni

* School of Human Evolution and Social Change, Arizona State University, PO Box 872402, Tempe, AZ 85287 Sarah.Mathew@asu.edu. Acknowledgements: Kimmo Eriksson and Pontus Strimling discussed the study topic with me and offered input on the study design. Robert Boyd, Joseph Henrich and Moshe Hoffman commented on the manuscript. Funding was provided the Swedish Research Council grant no. 2009-2390 and the John Templeton Foundation grant no. 48952

and Gachter, 2008). In contrast when pool punishment is allowed wherein punishment is meted by a single central entity paid for by group members, subjects more reliably resolve the second-order dilemma (Zhang et al., 2013; Hilbe et al., 2014; Baldassarri and Grossman, 2011). This has led to an emerging view that responsible second-order peer punishment does not exist or work and that peer sanctioning reflects a state of anarchy (Zhang et al., 2013). Moreover, compared to second-order free riders, punishers are liked less and are perceived to be aggressive rather than as other-regarding, which is surprising if punishment is a form of cooperation (Strimling and Eriksson, 2014).

A meaningful response to second-order free riding however may be hard to detect in societies with centralized punishment institutions that discourage peer sanctioning. If so, a society like the Turkana provide a uniquely suitable context to investigate second-order free riding and punishment. The Turkana are a Nilo-Saharan ethnolinguistic group in northern Kenya comprising about a million people. They periodically mobilize a few hundred warriors for cattle-raids of neighboring ethnic groups in which participants risk a 1 % chance of dying. Cowards and deserters are criticized by community members and may be subjected to corporal punishment and fines meted by their age mates—closely bonded groups of men born within a 4–7 year period who keep company with each other during communal gatherings, raids and while herding.

I presented Turkana adults with hypothetical vignette scenarios (see Appendix) in which I systematically varied how and whether a fictitious warrior who observes cowardice on a raid punishes the coward. After each scenario participants were asked questions to elicit their judgment of and motivation to reward or sanction the character. I also conducted a similar study in the context of adultery, a pairwise dispute, for which clan members of the offended man may impose corporal punishment on the violator and seize his animals. In another study I examined reaction to counter punishment, where a fictitious warrior retaliates against either a deserved or undeserved punishment for cowardice.

Compared to a character who administered appropriate punishment for cowardice (fig. 1a), participants were more likely to disapprove of a second-order free rider (Paired t-test: $t = 24.92, df = 488, \text{p-value} < .0001$), of a character who administers excess punishment (Welch Two Sample t-test: $t = 34.22, df = 709.12, \text{p-value} < .0001$), and of a character who falsely accused and punished someone of cowardice ($t = 42.2, df = 798.56, \text{p-value} < .0001$). Compared to second-order free riding, disapproval was higher for disbursing severe punishment (Welch Two Sample t-test, $t = 4.55, df = 907.16, \text{p-value} < .0001$) or wrongful punishment ($t = 7.38, df = 870.09, \text{p-value} < .0001$). In the context of adultery too (fig. 1b), subjects were more likely to disapprove of the conduct of a second-order free rider than they were of an individual who initiated appropriate sanctions (Paired t-test: $t = 31.82, df = 409, \text{p-value} < .0001$). Disapproval of counter punishment (fig. 1c) was higher when it was directed against an individual who disbursed appropriate punishment than when it was disbursed

against an individual who wrongfully punished the counter punisher (Paired t-test: $t = 14.73, df = 252$, p-value $< .0001$). Additionally, participants were more likely to judge punishment as wrongful if it is dispensed single-handedly without consulting others, or if it is meted by members outside the preordained group responsible for punishing (see fig A1, and Tables 4.1 and 4.2 in Appendix).

The results indicate that the Turkana have culturally evolved norms that help solve the second-order free rider problem and that promote group-beneficial punitive behavior. Compared to a character who metes responsible sanctions, a second-order free rider and individuals who punish excessively or unjustly elicit moral punitive reactions. Counter punishment is condoned when it is aimed at predatory punishment but is frowned upon when directed against legitimate punishment. Only by meting measured, responsible punishment can a person avoid disapproval.

Collective consensus from a preordained group of peers is important for punishment to be perceived as legitimate. Collective punishment is considered a means to lower the cost borne by each punisher (Boyd, Gintis and Bowles, 2010). But the fact that solo punishers are viewed not as exemplars but as wrongdoers implies that punishment is collective to ensure its moral legitimacy rather than for expediency. Consistent with Eriksson, Strimling and Ehn (2013), violations of norms restricting punishment elicited stronger disapproval than violation of norms prescribing punishment, but a confound here is that the the restrictive violation is an unauthorized act of violence whereas the prescriptive violation is merely an inaction. Oddly, meta-norms of adultery, a pairwise dispute, are similar to the meta-norms regarding free riding in warfare. Because clan members contribute to the brideprice for a man to marry, it may be that adultery is perceived as a violation that affects the clan and not just an individual.

Although there are evident norms condemning second-order free riding, there were no reported incidents in which a second-order free rider was directly sanctioned. Observable, unambiguous second-order free riding may be rare or hard to detect in real life preventing the opportunity for sanctioning. Alternately, since it is less costly to comply with a punishment norm than a first-order norm, subtler indirect consequences may suffice to maintain compliance.

Contra to the pattern emerging from study populations with centralized institutions, peer sanctioning among the Turkana is not anarchic, and is highly regulated even while conferring on every person an obligation and role in administering sanctions. Perhaps with centralized sanctioning the selection pressure on informal meta-norms peters out, causing group beneficial meta-norms to degenerate or be suppressed. This could be why antisocial punishment is common in existing lab studies of second-order punishment.

The findings highlight a way for norm enforcement to evolve. Among Turkana subjects the moral punitive psychology is evoked by second-order free riding in much the same way as it is by first-order free riding (see Mathew and Boyd (2014)). As is the case in formal judicial systems, a single machinery therefore

serves the tasks required to enforce a norm—compliance, sanctioning of violators, and disapproval of those who give violators a free pass. This means that if norms can arise about enforcement itself, the third- or higher-order dilemma can be obviated in informal punishment. Moreover processes like cultural group selection can yield moralistic punishment even without meta-norms as long as selection is sufficiently strong (Henrich and Boyd, 2001). The prevalence of group-beneficial meta-norms like those that exist in Turkana society will make it considerably easier for moralistic punishment to evolve via such processes.

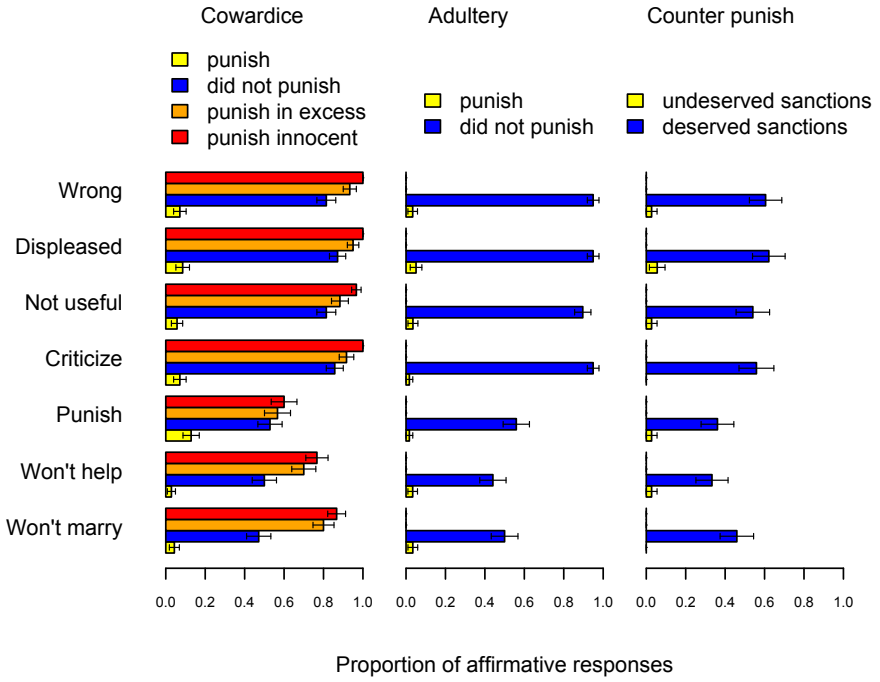

Figure 1. Peer punishment norms. Horizontal axis is the proportion of subjects giving affirmative responses to questions assessing disapproval of the vignette character. Left plot is reaction to a character who observes cowardice on a raid based on if he punished the coward or not ($n = 70$) and on whether sanctions were measured and deserved ($n = 60$). Center plot is reaction to a character who's spouse commits adultery based on if he sanctions or not (n =59). Right plot shows reaction to a fictitious warrior who retaliates against deserved or undeserved sanctions for presumed cowardice ($n = 38$). Error bars show 95 % CI.

REFERENCES

Baldassarri, Delia, and Guy Grossman. 2011. "Centralized sanctioning and

legitimate authority promote cooperation in humans." *Proceedings of the National Academy of Sciences of the United States of America*, 108(27): 11023–7.

Barclay, Pat. 2006. "Reputational benefits for altruistic punishment." *Evolution and Human Behavior*, 27(5): 325–344.

Boyd, R, H Gintis, and S Bowles. 2010. "Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare." *Science*, 328: 617–620.

Cinyabuguma, Matthias, Talbot Page, and Louis Putterman. 2006. "Can second-order punishment deter perverse punishment?" *Experimental Economics*, 9(3): 265–279.

Eriksson, Kimmo, Pontus Strimling, and Micael Ehn. 2013. "Ubiquity and efficiency of restrictions on informal punishment rights." *Journal of Evolutionary Psychology*, 11(1): 17–34.

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

Herrmann, Benedikt, Christian Thoni, and Simon Gachter. 2008. "Antisocial punishment across societies." *Science*, 319: 1362–1367.

Hilbe, Christian, Arne Traulsen, Torsten Röhl, and Manfred Milinski. 2014. "Democratic decisions establish stable authorities that overcome the paradox of second-order punishment." *Proceedings of the National Academy of Sciences of the United States of America*, 111(2): 752–6.

Kameda, Tatsuya, Masanori Takezawa, and Reid Hastie. 2003. "The Logic of Social Sharing: An Evolutionary Game Analysis of Adaptive Norm Development." *Personality and Social Psychology Review*, 7(1): 2–19.

Mathew, S, and R Boyd. 2011. "Punishment sustains large-scale cooperation in prestate warfare." *Proceedings of the National Academy of Sciences*, 108(28): 1091–6490.

Mathew, S, and R Boyd. 2014. "The cost of cowardice: punitive sentiments towards free riders in Turkana raids." *Evolution and Human Behavior*, 35(1): 58–64.

Strimling, Pontus, and Kimmo Eriksson. 2014. "Regulating the regulation: Norms about punishment." In *Social dilemmas: New perspectives on reward and punishment.*. , ed. P. A. M. Van Lange, B. Rockenbach and T. Yamagishi. New York:Oxford University Press.

Yamagishi, T. 1986. "The provisioning of sanctioning as a public good." *Journal of Personality and Social Psychology*, 51: 100–116.

Zhang, Boyu, Cong Li, Hannelore Silva, Peter Bednarik, and Karl Sigmund. 2013. "The evolution of sanctioning institutions: an experimental approach to the social contract." *Experimental Economics*.