Communicative & Integrative Biology 5:2, 1–5; March/April 2012; © 2012 Landes Bioscience

Could shame and honor save cooperation?

Jennifer Jacquet,[1,2,*] Christoph Hauert,[1] Arne Traulsen[3] and Manfred Milinski[4]

[1]Department of Mathematics; University of British Columbia; Vancouver, BC Canada; [2]Sea Around Us Project; University of British Columbia; Vancouver, BC Canada; [3]Research Group for Evolutionary Theory; Max-Planck-Institute for Evolutionary Biology; Plön, Germany; [4]Department of Evolutionary Ecology; Max-Planck-Institute for Evolutionary Biology; Plön, Germany

Shame and honor are mechanisms that expose behavior that falls outside the social norm. With recent six-player public goods experiments, we demonstrated that the threat of shame or the promise of honor led to increased cooperation. Participants were told in advance that after ten rounds two participants would be asked to come forward and write their names on the board in front of the fellow group members. In the shame treatment, the least cooperative players were exposed and wrote their names under the sentence "I donated least" while the honored participants wrote their name under "I donated most." In both the shame and honor treatments, participants contributed approximately 50% more to the public good, as compared with the control treatment in which all players retained their anonymity. Here, we also discuss how shame and honor differ from full transparency, and some of the challenges to understanding how anonymity and exposure modify behavior.

Keywords: cooperation, honor, public goods game, shame, tragedy of the commons

Submitted: 12/09/11

Accepted: 12/12/11

http://dx.doi.org/10.4161/cib.5.2.19016

*Correspondence to: Jennifer Jacquet; Email: jenniferjacquet@gmail.com

Addendum to: Jacquet J, Hauert C, Traulsen A, Milinski M. Shame and honour drive cooperation. Biol Lett 2011; 7:899–901; PMID:21632623; http://dx.doi.org/10.1098/rsbl.2011.0367

Can shame and honor lead to greater cooperation? We recently tested this question with a public goods game, a type of experiment often used to study the conflict between group and self-interest.[1,2] Over 12 rounds, six participants recruited from the same class could choose to invest a certain amount of their allotted capital, endowed to them at the start of the game, into a common pool or keep it. The total investments in the common pool were then doubled and equally divided among all participants. Thus, in each game non-contributors are better off than cooperators that contributed to the common pool. However, a group of non-contributors does not increase their initial capital, whereas a group of cooperators doubles their income. This generates a social dilemma characterized by the conflict of interest between the individual and the group.[3,4] This conflict of interest was beautifully illustrated by one student in a questionnaire handed out at the end of the experiments (**Fig. 1**).

In the shame treatment, we exposed, as we had announced at the start, the two participants who were least generous after the tenth round and, in the honor treatment, the two most generous players were revealed. These two players were asked to come to the front and write their real name on a board visible for all under the sentence "I donated least" (shame) or "I donated most" (honor). The other four players remained anonymous in both groups. Contributions to the public good were approximately 50% higher as compared with the control, where all six players knew that they would remain anonymous over all 12 rounds (**Fig. 2**). This demonstrates that both shame and honor can drive cooperation.

'Shame' and 'honor' are polysemous such that they can be used as nouns as well as verbs. In our experiments, we did not examine whether the participants felt the emotions shame or honor when their identity was revealed to the group, but rather whether they would respond differently to a cooperative experiment when there was a threat or an opportunity of being singled out in response to their actions. We employed shame and honor as verbs, in that we exposed individuals that made aberrant choices—in one case uncooperative relative to the group average

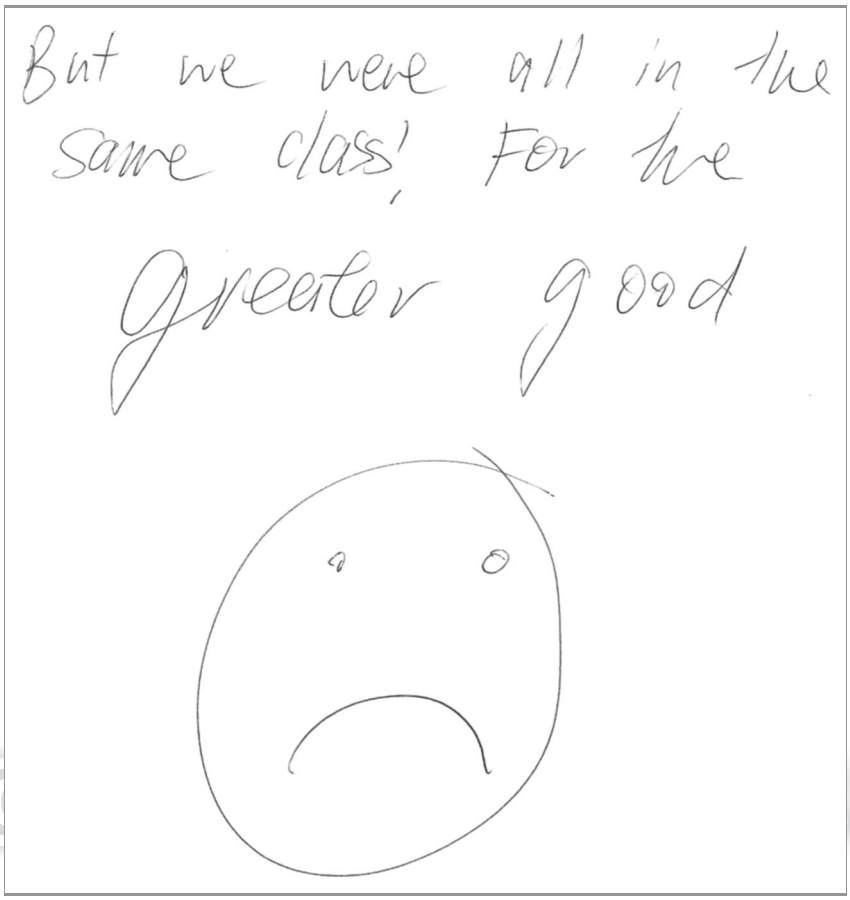

Figure 1. Feedback of a disillusioned student in the anonymous control treatment. The six participants were deliberatly recruited from the same class and early during the term so everyone knew that they would encounter one another again.

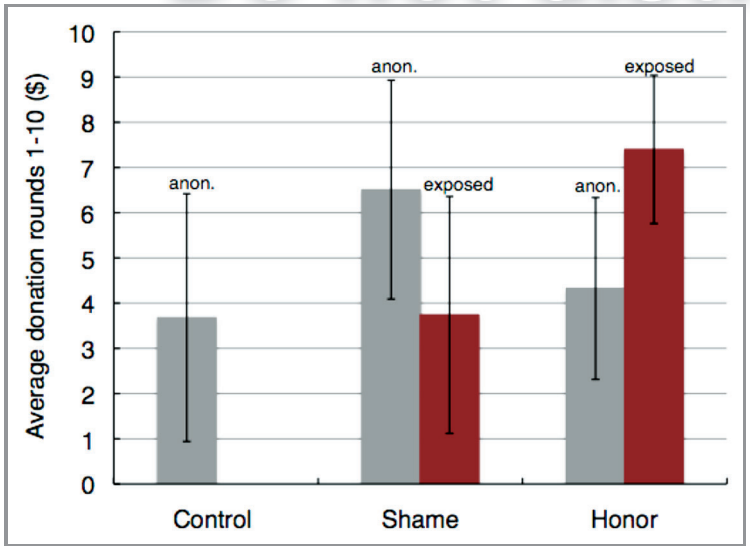

Figure 2. Average player contributions (and standard deviation) over the first ten rounds by treatment and player type (remaining anonymous or exposed after round 10; it had been made clear at the start that this would happen); maximum possible contribution was $10. In the control, all players were anonymous, while in shame and honor, four players remained anonymous and two players were exposed after round 10. Group contributions over the first ten rounds were significantly higher in the honor and shame treatments, as compared with the control.

and, in the other case, highly cooperative relative to the group average.

When it comes to shame and honor, the reference scale is provided by the currently accepted social norms. For example, the fear of being shamed for not adhering to standards for personal hygiene might encourage adherence to a social norm without the fear of physical punishment, fines, or incarceration. Shame and honor can be considered sub-classifications of punishment and reward. Whitman[5] sorts punishments into five sorts of deprivations: life (execution); liberty (imprisonment); bodily safety (corporal violence); property (fines); and what he calls "dignity" (shaming). We would also consider two more potential deprivations: access to the group (ostracism) and manipulation of social status (gossip). Shaming, ostracism, and gossip are often non-regulatory in the sense that members of the public can legally exert these punishments, while the law often consigns the use of execution, imprisonment, corporal violence, and fines as exclusive punishments of the state.

From an evolutionary perspective the emergence of any kind of punishment behavior that is costly to the punisher remains unclear. As long as such punishers are rare, they have to punish frequently and widely in order to enforce compliance and hence will perform poorly. Nevertheless, punishment is ubiquitous in human and animal societies.[6] Interestingly, theoretical and experimental work suggests that voluntary participation in social enterprises—which corresponds to the positive complement of ostracism—could be key, following Hardin's[2] principle of 'mutual coercion mutually agreed upon'. This principle has been demonstrated in recent theoretical work, where punishment behavior and sanctioning institutions can evolve if individuals voluntarily commit to such a social norm[7,8] and in experiments, where participants prefer—after an initial aversion and some adjustment time—public goods interactions that allow punishment of other participants.[9,10] Shaming and ostracism are two economic ways to inflict potentially severe and harmful sanctions on individuals.

Among pairs, cooperation can be maintained in repeated interactions through

direct reciprocity[11] but this becomes challenging in larger groups and fails if interaction partners change. Here, reputation and gossip play a decisive role in establishing and maintaining cooperation.[12] In particular, cooperation thrives if individuals gain reputation through cooperative actions and target their cooperative efforts toward others that have a high reputation. To learn the reputation of potential interaction partners, the spread of information or gossip becomes essential.[13] Similarly, gossip about whether an individual punishes non-contributors is predicted to help to enforce social norms and to stabilize cooperation.[14] The efficient interplay between reputation and punishment to promote cooperation has been demonstrated in an experimental setup.[10]

In real world settings, non-regulatory mechanisms are often of interest because state regulations can be difficult to instate and costly to enforce. Transparency policies, which reveal the behavior of every entity involved, are gaining popularity.[15] These include initiatives such as the public health policies in Los Angeles County and New York City where all restaurants must display a grade card (A, B, or C) that signals the restaurant's most recent government hygiene inspection (restaurants that fall below the 'C' score are closed).

Transparency requires that consumers of the information assess its reliability and set the thresholds for acceptable performance themselves, which can be quite an onerous task in times of an information overload. In contrast, policies that aim to shame or honor expose only a section of the population, such as the delinquent tax websites in many US states, and the threshold for acceptable behavior is often implicit. However, with shame and honor it can still be difficult to determine at which point behavior becomes aberrant. For taxpayers in arrears, for instance, some US states publish the names of only the top 250 delinquent taxpayers or those owing more than a certain amount (e.g., $100,000). In our own experiment, one group in the shame treatment was highly cooperative: five players fully cooperated and gave $10 over 10 rounds and one player gave $9. In this case, we tweaked the game, and at the end of 10 rounds exposed only the one player who gave $9. Moreover, in this case, we questioned whether the reference point of the group's average to determine what was 'shameful' was useful—shame and honor are probably most efficient when there is wide variance in behavior.

At least three previous experiments tested the outcome of transparency on cooperation by revealing the identity of all participants with variable degrees of exposure. In a one-round public goods game, every player had to reveal his or her contribution in front of the ten-person group and contributions increased from 34.4% of possible donations in the anonymous condition to 68.2%,[16] or a calculated 98% increase. In another eight-round game, players could see only digital photographs of all five group members along with their donations at each round and contributions increased from 30.3% of possible donations in an anonymous treatment to 48.1%,[17] or a calculated 60%. In a third design, the group of four players could talk before and after the experiments but were anonymous to each other until the tenth and final round, when all players' identities were revealed; group contributions increased from 50.5% of total possible contributions under anonymous conditions to 70.5%,[18] which translates to a 40% increase.

These studies reveal a trend of lower contributions with higher levels of anonymity, and higher contributions to the public good with lower levels of anonymity. In comparison, our experiments exposed only two out of six players and achieved a similar increase in cooperation. Groups threatened with shame and enticed with honor donated 53% and 48% more, respectively, than groups who could rely on retaining their anonymity. While the players 'honored' after round 10 gave significantly more than their co-players that remained anonymous (2 sample t-test, $n_{h(exposed)} = 20$, $n_{h(anon)} = 40$, $t = 5.91$, $p < 0.0001$; the player is the statistical unit; all statistical tests are 2-tailed; **Figure 2**), and the later 'shamed' players gave significantly less (2 sample t-test, $n_{s(exposed)} = 19$, $n_{s(anon)} = 41$, $t = -4.03$, $p < 0.0017$; **Figure 2**), the overall differences over the course of the first 10 rounds cannot be accounted for by the increased contributions of those participants that lost their anonymity and hence suggests that the patterns observed are not necessarily driven by the effects of publicity on low- or high-contributing participants. Instead, all players cooperated more.

For instance, the players who remained anonymous in the shame treatment gave more over 10 rounds than the top four most generous players in each group of the control treatment (2 sample t-test, $n_{s(anon)} = 41$, $n_{c(top4)} = 40$, $t = 2.92$, $p = 0.0046$). Overall, the distributions of

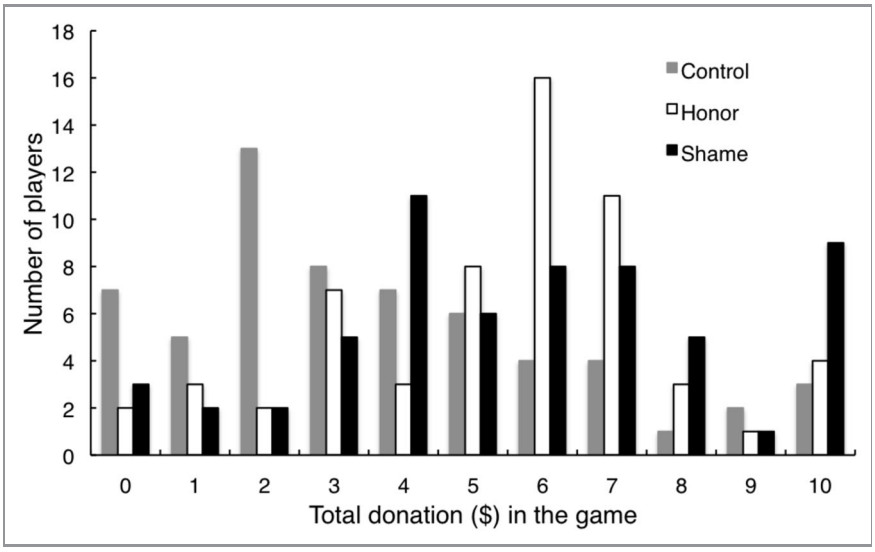

Figure 3. Histogram of the donations over first ten rounds of the experiment.

the numbers of players giving higher amounts shifted with the threat of shame or honor (**Fig. 3**). However, to clearly identify differences between the effects of transparency, anonymity, shame, and honor further experimental work is required. For instance, would cooperation also increase if groups were told two random players of the six would be exposed with their contribution rank? This set-up would demonstrate whether (and to what degree) cooperation increases due to the threat of mere exposure, as opposed to the threat of exposure for extreme behavior, as in our experiment.

Evidence from our experiment suggests that certain participants were responding to the specific threat of shame or the promise of honor. All subjects in our experiment also received a one-question survey following the experiment that asked: "What was your strategy when you decided to give or not in each round?" The answers also show that some participants made decisions arbitrarily. Responses included:

- "Eventually I just wanted to be known as the top contributor."
- "My strategy was to donate as little as possible without being exposed as someone who contributed least."
- "I did not want to be one of the 'the least generous players', so my only aim was to stay out of the bottom 2, other than that I tried to maximize profit."
- "Towards the 5th-6th rounds, my trend of thought changed, and I started paying attention to the individual contributions to make sure I was not in the bottom 2."
- "Give only in even rounds, plus the lucky number 7."
- "If I pulled out a coin and it was heads I donated. If it was tails, I wouldn't donate."

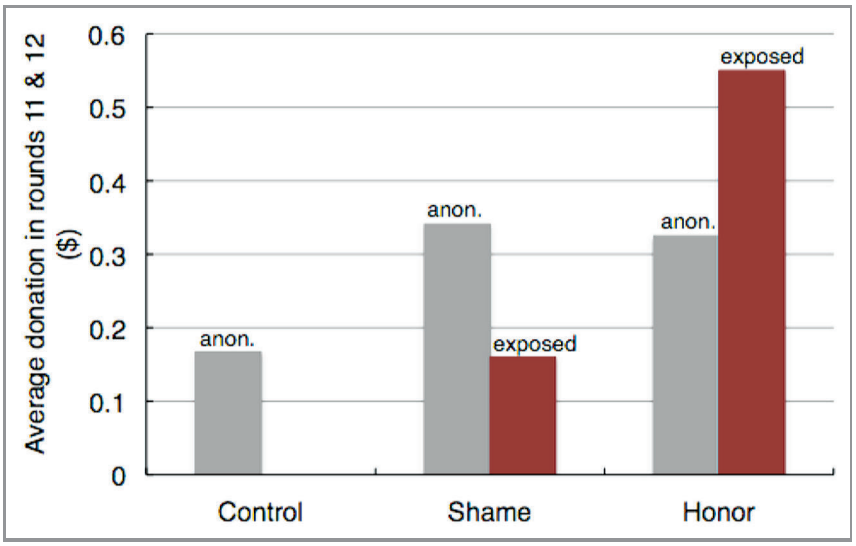

Figure 4. Average player contributions for rounds 11 and 12 by treatment and player type (remained anonymous or exposed after round 10); maximum possible contribution was $2. In the control, all players remained anonymous, while in shame and honor, four players remained anonymous and two players were exposed after round 10.

Questions remain about the potency of shame and honor. We exposed one-third of the group. As shame and honor are diluted (i.e., a greater proportion of the population exposed, which gets closer to full transparency) or made more acute (i.e., a smaller proportion), does this change their effectiveness? Many questions also remain related to the long-term consequences of shame and honor. We ran two additional rounds after exposure (**Fig. 4**) and saw that anonymous players in both the shame and honor treatments gave nearly equal amounts, while players who were honored contributed more in the final two rounds (and significantly more in the 12th round) than those who were shamed, suggesting that those who earned an honorable reputation felt obliged to maintain it.

Acknowledgments

This work was made possible with the help of the students of University of British Columbia and the funding of NSERC.

References

1. Gordon HS. The economic theory of a common property resource: the fishery. J Polit Econ 1954; 62:124-42; http://dx.doi.org/10.1086/257497
2. Hardin G. The tragedy of the commons. Science 1968; 162:1243-8; http://dx.doi.org/10.1126/science.162.3859.1243
3. Dawes R. Social dilemmas. Annu Rev Psychol 1980; 31:169-93; http://dx.doi.org/10.1146/annurev.ps.31.020180.001125
4. Hauert C, Michor F, Nowak MA, Doebeli M. Synergy and discounting of cooperation in social dilemmas. J Theor Biol 2006; 239:195-202; PMID:16242728; http://dx.doi.org/10.1016/j.jtbi.2005.08.040
5. Whitman J. What is wrong with inflicting shame sanctions? Yale Law J 1998; 107:1055-92; http://dx.doi.org/10.2307/797205
6. Clutton-Brock TH, Parker GA. Punishment in animal societies. Nature 1995; 373:209-16; PMID:7816134; http://dx.doi.org/10.1038/373209a0
7. Hauert C, Traulsen A, Brandt H, Nowak MA, Sigmund K. Via freedom to coercion: the emergence of costly punishment. Science 2007; 316:1905-7; PMID:17600218; http://dx.doi.org/10.1126/science.1141588
8. Sigmund K, De Silva H, Traulsen A, Hauert C. Social learning promotes institutions for governing the commons. Nature 2010; 466:861-3; PMID:20631710; http://dx.doi.org/10.1038/nature09203
9. Gürerk Ö, Irlenbusch B, Rockenbach B. The competitive advantage of sanctioning institutions. Science 2006; 312:108-11; PMID:16601192; http://dx.doi.org/10.1126/science.1123633
10. Rockenbach B, Milinski M. The efficient interaction of indirect reciprocity and costly punishment. Nature 2006; 444:718-23; PMID:17151660; http://dx.doi.org/10.1038/nature05229
11. Trivers R. The evolution of reciprocal altruism. Q Rev Biol 1971; 46:35-57; http://dx.doi.org/10.1086/406755
12. Nowak MA, Sigmund K. Evolution of indirect reciprocity by image scoring. Nature 1998; 393:573-7; PMID:9634232; http://dx.doi.org/10.1038/31225
13. Sommerfeld RD, Krambeck HJ, Semmann D, Milinski M. Gossip as an alternative for direct observation in games of indirect reciprocity. Proc Natl Acad Sci U S A 2007; 104:17435-40; PMID:17947384; http://dx.doi.org/10.1073/pnas.0704598104
14. Sigmund K, Hauert C, Nowak MA. Reward and punishment. Proc Natl Acad Sci U S A 2001; 98:10757-62; PMID:11553811; http://dx.doi.org/10.1073/pnas.161155698
15. Fung A, Graham M, Weil D. Full Disclosure: The Perils and Promise of Transparency, Cambridge, Massachusetts: Cambridge University Press, 2007.
16. Rege M, Telle K. The impact of social approval and framing on cooperation in public good situations. J Public Econ 2004; 88:1625-44; http://dx.doi.org/10.1016/S0047-2727(03)00021-5

17. Andreoni J, Petrie R. Public goods experiments without con□dentiality: a glimpse into fund-raising. J Public Econ 2004; 88:1605-23; http://dx.doi.org/10.1016/S0047-2727(03)00040-9

18. Gächter S, Fehr E. Collective action as a social exchange. J Econ Behav Organ 1999; 39:341-69; http://dx.doi.org/10.1016/S0167-2681(99)00045-1