

Each person sat at a computer in a cubicle, so nobody knew who their partners were on any particular round, although they saw a feedback screen after each round revealing exactly how much each of the four players had contributed. Also, after each round, Fehr and Gächter scrambled the groups so that each person played with three new partners—there was no chance to develop norms of trust, and no chance for anyone to use tit for tat (by holding back on the next round if anyone “cheated” on the current round).

Under these circumstances, the right choice for *Homo economicus* is clear: contribute nothing, ever. Yet in fact the students did contribute to the common pot—about ten tokens on the first round. As the game went on, however, people felt burned by the low contributions of some of their partners, and contributions dropped steadily, down to about six tokens on the sixth round.

That pattern—partial but declining cooperation—has been reported before. But here’s the reason this is such a brilliant study: After the sixth round, the experimenters told subjects that there was a new rule: After learning how much each of your partners contributed on each round, you now would have the option of paying, with your own tokens, to *punish* specific other players. Every token you paid to punish would take three tokens away from the player you punished.

For *Homo economicus*, the right course of action is once again perfectly clear: never pay to punish, because you will never again play with those three partners, so there is no chance to benefit from reciprocity or from gaining a tough reputation. Yet remarkably, *84 percent of subjects paid to punish*, at least once. And even more remarkably, *cooperation skyrocketed* on the very first round where punishment was allowed, and it kept on climbing. By the twelfth round, the average contribution was fifteen tokens.⁴⁶ Punishing bad behavior

promotes virtue and benefits the group. And just as Glaucon argued in his ring of Gyges example, when the threat of punishment is removed, people behave selfishly.

Why did most players pay to punish? In part, because it felt good to do so.⁴⁷ We hate to see people take without giving. We want to see cheaters and slackers “get what’s coming to them.” We want the law of karma to run its course, and we’re willing to help enforce it.

When people trade favors, both parties end up equal, more or less, and so it is easy to think (as I had) that reciprocal altruism was the source of moral intuitions about equality. But egalitarianism seems to be rooted more in the hatred of domination than in the love of equality per se.⁴⁸ The feeling of being dominated or oppressed by a bully is very different from the feeling of being cheated in an exchange of goods or favors.

Once my team at YourMorals.org had identified Liberty/oppression as a (provisionally) separate sixth foundation, we began to notice that in our data, concerns about political equality were related to a dislike of oppression and a concern for victims, not a desire for reciprocity.⁴⁹ And if the love of political equality rests on the Liberty/oppression and Care/harm foundations rather than the Fairness/cheating foundation, then the Fairness foundation no longer has a split personality; it’s no longer about equality *and* proportionality. It is primarily about proportionality.

When people work together on a task, they generally want to see the hardest workers get the largest gains.⁵⁰ People often want equality of outcomes, but that is because it is so often the case that people’s inputs were equal. When people divide up money, or any other kind of reward, equality is just a special case of the broader principle of proportionality. When a few members of a group contributed far more than the others—or,