

## Spite and Development

By ERNST FEHR, KARLA HOFF, AND MAYURESH KSHETRAMADE\*

*The disparity in the performance of economies and the persistence of disparate economies through time have not been satisfactorily explained by development economists.... what has been missing is an understanding of the nature of human coordination and cooperation.*

— Douglass North (1990, 11)

Effective institutions for contract enforcement and collective action are probably among the most important conditions for successful economic and social development. However, in many developing countries these institutions are weak and agreements lacking formal third-party enforcement are ubiquitous. Under these conditions, endogenous contract enforcement and the ability of private parties to solve the cooperation and coordination problems inherent in collective action are key. A large body of experimental evidence suggests that social preferences are vital to solving such problems.<sup>1</sup> In particular, the willingness to punish cheaters in informal agreements, even at a net cost to the punisher, substantially reduces cheating (Fehr, Gächter, and Georg Kirchsteiger 1997), and the willingness to altruistically punish free riders in public goods greatly enhances the scope of private parties to solve collective action problems (Fehr and Gächter 2002). Likewise, many people exhibit a propensity to cooperate conditional on others' cooperation, even when free-riding

would maximize their material payoff (Urs Fischbacher, Gächter, and Fehr 2001).

In this paper, however, we document that social preferences may also constitute important obstacles to development. We show that spiteful preferences—the desire to reduce another's material payoff for the *mere* purpose of increasing one's *relative* payoff—are surprisingly widespread in experiments we conducted in one of the least developed regions in India (Uttar Pradesh). Our results suggest that spitefulness diminishes the propensity to cooperate, even in those situations in which mutual cooperation is an equilibrium for selfish players. Finally, we find empirical patterns suggesting that the willingness to reduce another's material payoff—either for the sake of achieving more equality or for the sake of being ahead—is stronger among individuals belonging to high castes than among those belonging to low castes.

The last finding is of particular interest because individuals' caste status can be considered as exogenous—individuals cannot freely select into different castes but inherit the caste status of their parents.<sup>2</sup> For this reason, and because we control for demographic variables such as education, land ownership, and house type, our finding may represent the impact of a caste culture that puts extreme emphasis on the superiority of the high castes and the inferiority of the low castes.<sup>3</sup> It seems quite plausible that such a culture contributes to social preferences such as the motive to be superior in terms of

\* Fehr: Institute for Empirical Research in Economics, University of Zurich, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland (e-mail: efehr@iew.uzh.ch); Hoff: World Bank, 1818 H Street, NW, Washington, DC 20433 (e-mail: khoff@worldbank.org); Kshetramade: Affinova, Inc., 52 Second Ave., Waltham, MA 02451 (e-mail: mayurv@yaho.com). We gratefully acknowledge support from the Research Priority Program on the "Foundations of Human Social Behavior" at the University of Zurich (Fehr) and from the World Bank (Hoff). We thank Sonal Vats for excellent research assistance.

<sup>1</sup> See Christopher M. Anderson and Louis Putterman 2006; Jeffrey Carpenter 2007; Fehr and Simon Gächter 2000; Ozgur Güreker, Bernd Irlenbusch, and Bettina Rockenbach 2006; and Elinor Ostrom, James Walker, and Roy Gardner 1992.

<sup>2</sup> Also, the assignment of different social groups and individuals to the two extreme ends of the caste hierarchy generally occurred so long ago (at least a millennium, possibly two) that any cultural differences across castes at the extreme ends of the caste hierarchy that may have determined that assignment should have by now been erased unless they were reinforced by the socioeconomic regime in which the caste groups have lived.

<sup>3</sup> The Indian caste system represents an extreme form of social hierarchy. High castes traditionally could command forced labor from low-caste individuals. Low castes are those groups that have been denied basic rights and were subject to the traditional practice of untouchability.

material payoff, even when this comes at a cost to equality and efficiency.

### I. The Role of Spite in Cooperation and Punishment

In symmetric public goods experiments, players have the same endowment and the same material payoff functions. Free-riders earn a higher material payoff than cooperators in this setting. If a punishment opportunity is introduced such that individuals are first informed about other group members' contributions to the public good, and can subsequently target punishment to specific individuals, many cooperators punish the free riders (Fehr and Gächter 2002). Such punishment occurs despite the fact that the punisher has to pay for sanctioning others. However, there is also evidence indicating that some free-riders punish the cooperators. For example, in a symmetric one-shot public goods experiment with a punishment opportunity (Armin Falk, Fehr, and Fischbacher 2005), roughly 13 percent of the subjects free ride and punish the cooperators. Interestingly, this punishment occurs only if the cost for the punisher is smaller than the cost for the punished subject, i.e., only if the free riders can increase the difference in material payoffs between themselves and the punished subjects. If instead, every \$1 invested in punishment reduces the punished subject's income by \$1, spiteful punishment completely vanishes. This finding is consistent with the view that free riders who punish want to increase the payoff difference between themselves and the cooperators.

Social psychologists also have found evidence for spiteful preferences (Paul A. M. Van Lange 1999). Van Lange examined the social preferences of more than 2,000 subjects in the Netherlands with the ring test (Wim B. G. Liebrand and Charles G. McClintock 1988) and found that roughly 12 to 13 percent of them are willing to pay for increasing inequality. These subjects prefer, for example, the allocation (480 for self, 80 for other) relative to the allocation (540 for self, 280 for other), thus sacrificing total surplus and equality for the sake of a larger payoff difference between "self" and "other."

In a wide variety of settings, spiteful preferences would constitute an obstacle to trade, cooperation, and, thus, development. A spiteful individual is willing to forgo material gains from trade unless the terms of trade give him a large

share of the pie. A spiteful individual is also more likely to violate contracts—either by providing low effort or low quality or by a lower willingness to pay the bill—because such contract violations increase the shirker's payoff at the expense of the trading partner. If spiteful shirking is anticipated, it will further decrease the willingness to trade. A spiteful individual is harder to motivate to cooperate because he has a higher marginal cost of contributing to public goods or joint activities: First, he bears the pecuniary costs like all other individuals and, second, he has a nonpecuniary cost because any contribution may reduce the difference in material payoffs between himself and others. Finally, spiteful punishment also has a detrimental impact on cooperation because it diminishes the *net* incentive for cooperation that emerges from the punishment of defectors. If defection *and* cooperation are punished, the potential free-riders have less incentive to cooperate. In fact, Gächter and Benedikt Hermann (2007) provide evidence from experiments in Russia indicating that punishment opportunities may fail to reduce free-riding because of the punishment of cooperators.

### II. Spiteful Punishment in India

We conducted a sequential, one-shot exchange game with third-party punishment in Uttar Pradesh in order to study the potential impact of an extreme social hierarchy on the willingness of disinterested third parties to punish violations of informal agreements. In our experiment, three players—A, B, and C—are involved and have an endowment of 50, 50, and 100 rupees, respectively. Fifty rupees are roughly equal to one day's unskilled wage. Players A and B participate in a binary trust game, with A in the role of the trustor and B in the role of the trustee. Player C is the third party who has the option of punishing B. A has to choose whether to keep his endowment (and then the game ends) or to send it to the trustee. If A sends his endowment to B, the experimenter triples the amount sent. At this point, B has 200 rupees. He has to choose whether to keep them for himself or to share them equally with the trustor. If B shares the money with A, all three players have 100 rupees before C makes a decision. If B keeps all the money, the distribution before C makes a decision is (0, 200, 100).

Player C has more than two options. For every two-rupee coin that C spends on punishment,

Player B loses a 10 rupee note. We asked C to make a choice for the case where B keeps all the money and also for the case where B shares the money with A. C makes this choice before he learns B's decision. This game captures a feature of many market exchanges and investment decisions, namely, that in order to obtain a social surplus, an individual exposes himself to the risk of opportunism. Since the game is played one-shot and the players who interact in the game are drawn from different villages whose names are never revealed, it is never in the self-interest of C to punish B. Thus, punishment represents *some kind of social preference*.

The rules of the game were explained to the subjects at great length. Subjects who did not pass a basic test of comprehension did not go on to participate in the game. Because we were interested in the role of caste status on the punishment of defection, we implemented four treatment conditions with a different composition of high (H) and low (L) castes, where the last names indicated a player's caste: HHH (all three players are from H), HLH (only B is from a low caste), LLL (all three players are from L), and LHL (only B is from a high caste). In total, 621 adult male players participated with 205 players in the role of player C. We find significant differences in altruistic punishment across treatment groups, which enable us to discriminate between the caste culture hypothesis, the caste conflict hypothesis, and the caste submission hypothesis.<sup>4</sup> These results are documented and discussed in depth in Hoff, Kshetramade, and Fehr (2007). In addition, we find a surprisingly large number of cases in which third parties punished cooperative Bs. In fact, across all four conditions, between 61 and 73 percent of the third parties punish cooperators, and none of the small treatment differences is significant (see Figure 1). On average, punishment of defectors is greater than punishment of cooperators

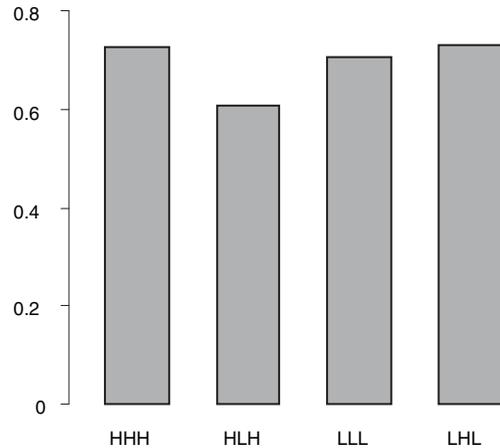


FIGURE 1. ANTISOCIAL PUNISHMENT BY THIRD PARTIES

in each of the four treatments. This does not, however, diminish the general importance of antisocial punishment because by decreasing the expected returns to cooperation for player B, such punishment diminishes the defection-detering force of altruistic punishment.

Subjects' verbal remarks after the experiment, when we asked them for their motives for punishing cooperators, indicate the spiteful nature of their punishment. We noticed very many statements such as "I want B to lose," "To reduce B compared to me," "I want to defeat B," "I am jealous of B; that is why it is important to impose a loss on him," "I want to bring down B," "Imposing a loss in the game gives enjoyment," "When somebody loses and goes home, it gives enjoyment," "I wanted to destroy B." The behavioral evidence, together with the verbal statements, provides clear evidence for the spitefulness involved in punishing cooperators.

### III. The Potential Impact of Spite on Cooperation and Coordination

In order to study the impact of spiteful preferences on cooperation and coordination, Hoff and Priyanka Pandey (2007) conducted a stag hunt game, in which both mutual cooperation and mutual defection are equilibria for selfish players (see Figure 2 in which the equilibrium cells are shaded). The mutual cooperation equilibrium is better for both players, but in the case of a unilateral defection, the cooperating player earns only 3, while the defecting player earns 7. Inequality aversion in the sense of Fehr

<sup>4</sup> The *caste conflict* hypothesis states that in the treatment with mixed caste groups, altruistic punishment will be higher than in the single caste group treatment (HLH > HHH and LHL > LLL). The *caste submission* hypothesis states that moral outrage and thus altruistic punishment is smaller if an H does not share with an L, compared to an L not sharing with an L (LHL < LLL). The *caste culture* hypothesis states that, compared to low-caste individuals, high-caste members are more disposed to punish unfair treatment because, among other factors, a long history of denial of basic rights means that low-caste members feel less entitled to fair treatment (HHH and HLH > LLL and LHL).

	Cooperate	Defect
Cooperate	10, 10	3, 7
Defect	7, 3	7, 7

FIGURE 2. THE STAG HUNT GAME

and Klaus M. Schmidt (1999), and spitefulness as defined here, will increase the likelihood of choosing the noncooperative action in this game. If a player dislikes disadvantageous inequality, the expected utility of cooperation is decreased (for any interior belief about the other player's action) since the outcome (3 for self, 7 for other) gives rise to a nonpecuniary disutility. In addition, if a player likes advantageous inequality, the expected utility of defection is increased (for any given interior belief about the other's behavior) since he derives extra utility from being ahead in case the other player cooperates and he defects. Thus, everything else equal, a spiteful player is more likely to choose the noncooperative action.

Subjects played the game in Figure 2 ten times, using tokens that at the end of the game were exchanged for an equal number of rupees. During the first five periods, subjects had one partner. After every period, they received feedback about their partner's choice. In periods 6 to 10, subjects were paired with another partner with whom they played the game again for five periods. Hoff and Pandey conducted three treatments that differ only in the caste composition of pairs: an LL treatment (both players from a low caste), an HH treatment, and an LH treatment. This design can address the question whether high-caste members are more or less able than low-caste members to coordinate on the good equilibrium, and whether mixed pairings do worse.

The results are that the high castes are much less able to coordinate on the good equilibrium—a difference across castes that is highly significant. In round 5, the LL pairs obtain the good equilibrium in 67 percent of the cases (10/15), whereas the HH pairs obtain it in only 19 percent of the cases (3/16). The HH pairs do even worse than the mixed LH pairs, who reach the good equilibrium in 40 percent of the cases (12/30). In the second half of the experiment, in which players change partners, coordination on

the good equilibrium increases in all treatments, and LL pairs achieve the Pareto-dominant outcome in 80 percent of the cases (12/15). The HH pairs are still the worst and achieve the good equilibrium in only in 47 percent of the cases (7/15).

The striking difference in cooperation across castes could be due to three main sources: (a) high-caste members could have a stronger aversion to disadvantageous inequality; (b) high-caste members could be more spiteful generally and more angry at "playing the fool"; (c) high-caste members could have less optimistic beliefs about their partner's willingness to cooperate. The "belief hypothesis" seems rather implausible because, even in the final round of a game with a given partner, that is, after the expectations had time to converge, the high-caste members cooperate much less. This interpretation is further supported by the fact that if the opponent cooperated in the first round, the high-caste members (compared to the low-caste members) are less likely to cooperate in the second round of the five-round game. However, because the stag hunt game does not allow us to make inferences about preferences, Hoff and Pandey conducted additional experiments to examine potential differences in preferences across castes.

#### IV. Are Members of High Castes More Spiteful?

Table 1 reports a series of binary choice dictator games that explore the prevalence of positive and negative inequity aversion, as well as of pure spite across castes. A total of 240 subjects participated in these games, each one of them making a choice in only one of the games. Each subject participates in one session and plays first the role of Player B (who is the dictator) in a game, and then the role of Player A (who is the recipient) in a different game. He does not learn that he will play the role of A until he has completed his role as B. Participants are divided into two groups in nearby locations. Every subject in one group is randomly matched with another subject in the other group; the matches are independent across the two games. In order to rule out any social pressure, the players' decisions were double-blind. The protocol made it transparent that no one, including the experimenter, could learn the decisions of any individual player. The only thing that the subjects know about their partner is his caste status (high or low) and that he is

TABLE 1—RESULTS FOR BINARY CHOICE DICTATOR GAMES

Game	N*	Payoffs (rupees for other, <b>rupees for self</b> ) for left option and right option	Fraction who choose left option		<i>p</i> -value	
			High caste	Low caste	Two-sided	One-sided
1	60	B chooses (90, <b>90</b> ) vs. (70, <b>100</b> )	0.40	0.47	0.609	
2	76	B chooses (70, <b>90</b> ) vs. (90, <b>90</b> )	0.42	0.21	0.049	0.024
3	90	B chooses (100, <b>100</b> ) vs. (180, <b>110</b> )	0.44	0.29	0.128	0.064
4	60	B chooses (100, <b>150</b> ) vs. (160, <b>160</b> )	0.30	0.33	0.785	0.40
5	60	B chooses (160, <b>100</b> ) vs. (110, <b>110</b> )	0.53	0.53	1.00	
6	74	B chooses (150, <b>120</b> ) vs. (100, <b>130</b> )	0.46	0.62	0.166	
7	60	B chooses (150, <b>150</b> ) vs. (100, <b>160</b> )	0.83	0.53	0.01	

Notes: Bold numbers indicate the dictator’s payoff. We report the total number of subjects. For each game, half the subjects are high caste and half are low caste. Thus, for example, in game 1, there are 30 high-caste subjects and 30 low-caste subjects. The two-sided *p*-value refers to the null hypothesis of no caste differences and is based on a t-test. Based on the results of the previous section, the null hypothesis for games 2 to 4 is that high castes are more spiteful or more willing to pay to reduce disadvantageous inequality; therefore, we added one-sided *p*-values for these games.

from their village. Hoff and Pandey matched all high-caste players with high-caste players, and all low-caste players with low-caste players because they were interested in how the dictators behave in HH and LL pairs.

For our purposes, game 2 is particularly interesting. In this game, B can decrease A’s payoff without cost to himself. If B does so, he reduces the total payoff and violates the equality norm. Forty-two percent of the high-caste members but only 21 percent of the low-caste members behave in this spiteful way—a difference that is significant. In game 4, player B has again a similar option, but this time behaving spitefully is costly for B. Here, we see no difference between the two castes, but it is still striking that roughly one-third of the players in both castes prefers the spiteful allocation. Game 3 is also interesting because here B can escape disadvantageous inequality by reducing A’s payoff. In this game, the high-caste players again reduce A’s payoff more often (in 44 percent versus 29 percent of the cases) but the difference is only weakly significant.

In games 5 to 7, B can sacrifice 10 rupees for the sake of increasing his partner’s payoff by 50 rupees. In games 5 and 6, sacrificing leaves the dictator with a lower payoff than his partner. In game 7, sacrificing leaves the dictator with the same payoff as his partner. The *high*-caste subject is more likely to sacrifice to help his partner in game 7 (83 percent are willing to sacrifice) than in games 5 and 6 (about 50 percent are willing to sacrifice). The *p*-value of the difference in proportions of game 7 versus games 6 and 5 is 0.06. In contrast, the fraction of the *low*-caste

subjects who are willing to sacrifice to help their partners is about the same across the three games (53 percent sacrifice in game 7, whereas on average 57 percent sacrifice in games 5 and 6). Logit regressions based on all data in Table 1 permit estimation of the “average” aversion against disadvantageous inequality and against advantageous inequality with caste-specific parameters.<sup>5</sup> The results indicate that both high- and low-caste subjects value the recipient’s payoff significantly positively if ahead ( $\beta > 0$ ), and that there are no significant caste differences. If, however, a dictator’s payoff is lower than the recipient’s payoff, the high-caste subjects value the recipient’s payoff highly and significantly negatively ( $\alpha < 0$ ), while the low-caste subjects even care positively for the other’s payoff ( $\alpha > 0$ ) although this parameter is not significant.

To summarize, these results show that, on average, high-caste subjects value others’ payoff strongly negatively if they experience disadvantageous disutility, and in some circumstances (game 2), the prevalence of spiteful preferences is much larger among high-caste subjects. Taken together, these results provide a plausible rationale for the lower ability of high-caste subjects to cooperate and coordinate on superior equilibria.

<sup>5</sup> In the spirit of Gary Charness and Matthew Rabin (2002), the following utility function over payoffs to A and B, denoted  $\pi_A$  and  $\pi_B$ , provides the basis for the logit estimates:  $U_B = [1 - \sigma] \pi_B + \sigma \pi_A$  if  $\pi_B < \pi_A$  and  $U_B = [1 - \rho] \pi_B + \rho \pi_A$  if  $\pi_B \geq \pi_A$ . This is equivalent to a Fehr-Schmidt utility function  $U_B = \pi_B - \alpha(\pi_A - \pi_B)$  if  $\pi_B < \pi_A$  and  $U_B = \pi_B - \beta(\pi_B - \pi_A)$  if  $\pi_B \geq \pi_A$  because  $\beta = \rho$  and  $\alpha = -\sigma$ .

## V. Concluding Remarks

Spiteful preferences may constitute a considerable obstacle for trade, cooperation, and, thus, development. In this short paper, we documented a surprisingly large amount of spiteful behavior in experiments conducted in India. A large majority of subjects punishes cooperative behavior in a one-shot trust game, although such punishment increases inequality and decreases the total payoff of both subjects. Evidence in dictator games suggests that high-caste subjects (compared to low-caste subjects) are considerably more likely to reduce others' payoffs if behind, or to take other spiteful actions. In addition, high-caste subjects are strikingly less capable of coordinating on welfare-improving equilibria. We believe that this lower ability to cooperate may be due to the high-caste subjects' concern for status and superiority and their strong aversion to disadvantageous inequality. Because extreme social hierarchies are typically accompanied by a culture that stresses status-seeking, it is a plausible hypothesis that the observed social preference patterns are at least partly shaped by this culture. Thus, an exciting question for future research is the extent to which different institutions and cultures produce preferences that are conducive or detrimental to economic development.

## REFERENCES

- ▶ **Anderson, Christopher M., and Louis Putterman.** 2006. "Do Non-Strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism." *Games and Economic Behavior*, 54(1): 1–24.
- ▶ **Carpenter, Jeffrey .** 2007. "The Demand for Punishment." *Journal of Economic Behavior and Organization*, 62(4): 522–42.
- ▶ **Charness, Gary, and Matthew Rabin.** 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117(3): 817–69.
- ▶ **Falk, Armin, Ernst Fehr, and Urs Fischbacher.** 2005. "Driving Forces Behind Informal Sanctions." *Econometrica*, 73(6): 2017–30.
- Fehr, Ernst, and Simon Gächter.** 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90(4): 980–94.
- ▶ **Fehr, Ernst, and Simon Gächter.** 2002. "Altruistic Punishment in Humans." *Nature*, 415(6868): 137–40.
- ▶ **Fehr, Ernst, and Klaus M. Schmidt.** 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817–68.
- ▶ **Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger.** 1997. "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica*, 65(4): 833–60.
- ▶ **Fischbacher, Urs, Simon Gächter, and Ernst Fehr.** 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71(3): 397–404.
- Gächter, Simon, and Benedikt Hermann.** 2007. "The Limits of Self-Governance When Cooperators Get Punished: Experimental Evidence from Urban and Rural Russia." The Centre for Decision Research and Experimental Economics Discussion Paper 2007–11.
- ▶ **Güerker, Özgür, Bernd Irlenbusch, and Bettina Rockenbach.** 2006. "The Competitive Advantage of Sanctioning Institutions." *Science*, 312(5770): 108–11.
- Hoff, Karla, and Priyanka Pandey.** 2007. "Understanding Obstacles to Efficient Coordination with Simple Tests: Experimental Evidence from India." Unpublished.
- Hoff, Karla, Mayuresh Kshetramade, and Ernst Fehr.** 2007. "Does Social Exclusion of a Group Undermine Its Trust-Enforcing Mechanisms? Evidence from a Third Party Punishment Experiment." Unpublished.
- ▶ **Liebrand, Wim B. G., and Charles G. McClintock.** 1988. "The Ring Measure of Social Values: A Computerized Procedure for Assessing Individual Differences in Information Processing and Social Value Orientation." *European Journal of Personality*, 2(3): 217–30.
- North, Douglass.** 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- ▶ **Ostrom, Elinor, James Walker, and Roy Gardner.** 1992. "Covenants with and without a Sword: Self-Governance Is Possible." *American Political Science Review*, 86(2): 404–17.
- ▶ **Van Lange, Paul A. M.** 1999. "The Pursuit of Joint Outcomes and Equality in Outcomes: An Integrative Model of Social Value Orientation." *Journal of Personality and Social Psychology*, 77(2): 337–49.