

What Can Economists Explain by Taking into Account People's Perceptions of Fairness?

Punishing Cheats, Bargaining Impasse, and Self-Perpetuating Inequalities

For some time economists have wrestled with the question whether, and in what domains of application, a theory of economic behavior should take into account individuals' perceptions of fairness (see, for example, Kahneman et al. 1986, Rabin 1993, and Fehr and Schmidt 2003). There is a standard hypothesis in economics, the *rational self-interest hypothesis*, which is based on a radically simplified view of human nature. In this view, individuals are exclusively motivated by their material self-interest and unboundedly rational in the pursuit of it. This hypothesis provides accurate predictions for competitive markets with standardized goods. However, much economic activity occurs outside of such markets—in markets with a small number of traders, within firms, under incompletely specified or incompletely enforceable contracts, or within neighborhoods through collective action. The experimental literature has been successful in showing why, in competitive environments with standardized goods, models based on the rational self-interest hypothesis provides accurate predictions even when individuals have strong preferences for fairness, and why in other environments, the self-interest models are refuted and alternative models in which individuals care about fairness and reciprocity are supported.

Experiments have played a critical role in the understanding of fairness-driven behavior and of the strategic settings in which preferences for fairness will matter.

Behavior in real-life circumstances can almost always be attributed to a wide set of possible motives. Therefore knowledge about specific motives predominantly stems from experiments—settings in which subjects interact under carefully controlled conditions, each of which can be varied individually in order to analyze its effect. In a typical experiment, subjects are randomly assigned to a treatment. Subjects do not know the personal identity of those with whom they interact and have full knowledge of the structure of the experiment. In all experiments conducted by economists, real money, sometimes substantial, is at stake.

The rest of this paper is organized as follows. Section 1 reviews experimental evidence that overwhelmingly refutes the hypothesis that individuals are concerned only with their material self-interest. This section also shows why in some competitive environments, the self-interest hypothesis is nevertheless so successful.

Section 2 shows that the incidence in the population of some individuals with preferences for being fairly treated and treating others fairly can solve a free rider problem. Whereas a deep historical root of modern economics is Adam Smith's insight that exclusive concern with self-interest can lead to the greatest good for the greatest number, there are many settings where concern for fairness—and the willingness at a cost to oneself to punish those who violate cooperation norms—is vital to achieving efficiency.

Section 3 demonstrates a pitfall of the human concern for fairness. Combined with the bounds on individuals' ability to process information and to reason in complex situations, fairness concerns can lead to bargaining impasse. Individuals tend to judge what is fair in an (unconsciously) self-interested way. Biased perceptions and judgments

may eliminate the possibility of reaching agreement. Fairness concerns in this way can be self-defeating.

Section 4 provides examples of pathways through which historical institutions that created extreme inequality across observationally distinct social groups can be perpetuated through belief systems that represented the oppression as “fair” because the dominated group was innately inferior.

1. When do preferences for fairness matter?

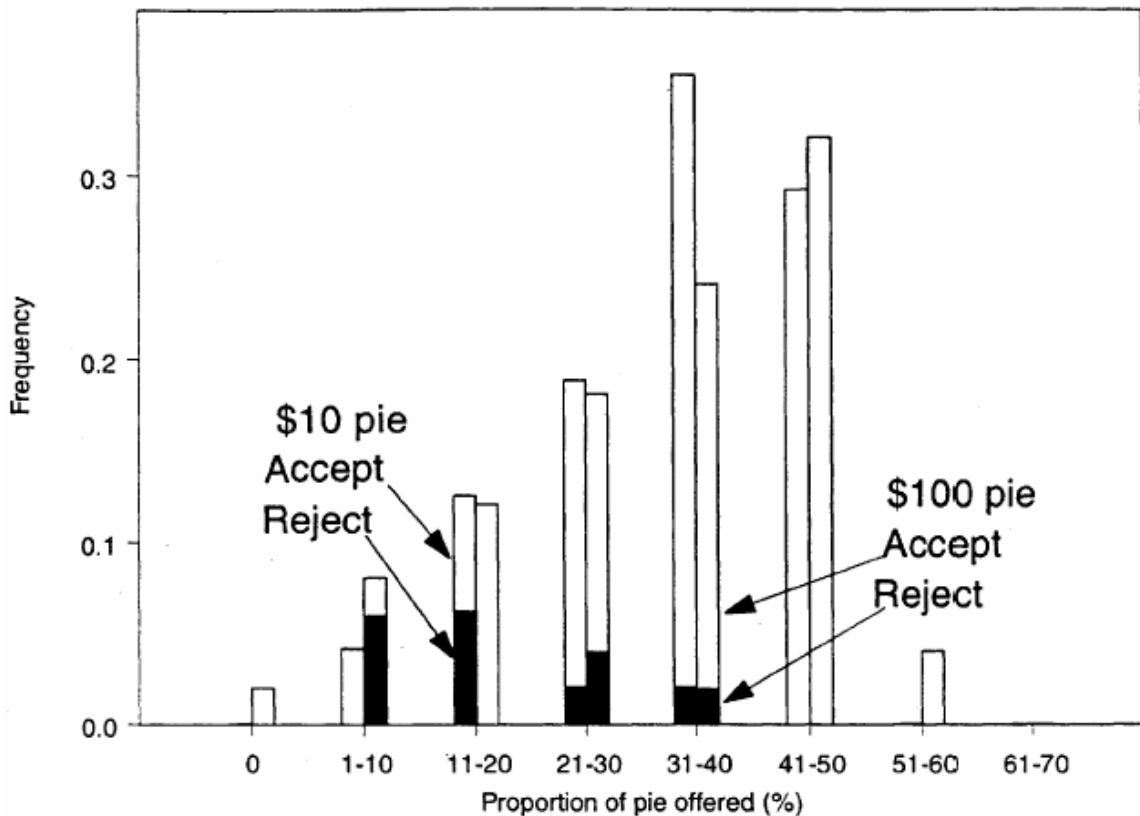
A two-player bargaining environment called the Ultimatum Game demonstrates that a sizeable number of people, from a wide variety of cultures, are willing to punish others at a cost to themselves to sanction unfair behavior (Guth, Schmittberger and Schwarze 1982, Thaler 1988 are seminal papers; a review of a vast literature is Camerer 2003). In this game, one bargainer makes a proposal of how to divide a certain sum of money with another bargainer, who has the opportunity to accept or reject the proposed division. The first bargainer is called the proposer, and the second is called the responder. If the responder accepts the proposed division, each bargainer earns the amount proposed, and if the responder rejects, then each bargainer earns zero.

The ultimatum game can be viewed as a model of what happens in the last minutes before a labor strike is called, or at the threshold of the courthouse door before a lawsuit goes to trial. That is, it is a model of the last step in the usual protracted process in which bargaining occurs. The game is interesting because it sheds light on three questions: Will the responder accept “unfair “ offers? What does the proposer believe about the responder’s acceptance strategy? Is the proposer motivated only by the threat

of rejection? By moving first, the proposer has all the bargaining power and can exploit it under the pure self-interest hypothesis, because under that hypothesis the responder will accept any positive amount of money. The rational self-interest hypothesis predicts that the proposer gives the responder the smallest money unit possible and keeps the rest.

The game has been run for stakes that range from a few dollars to a few months of income in many countries. Yet the pattern of findings is relatively constant. (See however in Section 4 a discussion of differences in some tribal cultures.) Across hundreds of experiments, proposals offering the responder less than 20 percent of the available surplus are rejected with probably .4-.6. The probability of rejection is decreasing in the size of the offer.

Figure 1: Offers and rejections in high-and low-stakes ultimatum games



Source: Based on data from Hoffman, McCabe, and Smith (1996)

Figure 1 shows the results from an experiment by Hoffman et al. (1996). The experiment was conducted in two ways: In the first treatment, the proposer and responder bargain over \$10; in the second treatment, they bargain over \$100. The figure shows that the same pattern of results occurs whether the total surplus is \$10 or \$100. The horizontal axis shows the amount offered to the responder, and the vertical axis shows the relative frequency of offers of different amounts. The dark part of each frequency bar is the fraction of offers that were rejected. The figure shows that the modal offer in the first treatment was between \$3 and \$4, and the modal offer in the second treatment was between \$41 and \$50. Offers ranged from 0 to 60 percent of the “pie,” most offers were close to half, and low offers were often rejected.

In the case of rejection, the responder is deciding that he wants both players to get nothing, rather than that he himself get a small share while the proposer gets much more. The Ultimatum Game results show that a large fraction of individuals has a preference for fairness and is willing to forego a benefit to punish behavior that they perceive as unfair, and also that most proposers seem to understand that low offers will be rejected. Therefore the equal split is often the modal offer. The central role of rejections is indicated by the Dictator Game, in which the proposer *unilaterally* dictates the division of the money because the responder cannot reject the offer. The average amount given to the responders in the Dictator Game is much less than in the Ultimatum Game (Forsythe et al. 1994).

A simple modification of the Ultimatum Game shows why the self-interest hypothesis nonetheless is very successful in explaining the outcome of competitive

markets with products of fixed quality. The modification introduces proposer competition into the Ultimatum Game. This game was conducted by Roth et al. (1991) in four cities: Jerusalem, Ljubljana (Slovenia), Pittsburgh, and Tokyo. Nine proposers each submit an offer to one responder. The responder can either accept or reject the highest offer. If the highest offer is made by several proposers, then one of them is selected at random. The offer by the successful proposer is implemented, whereas all the other proposers get 0.

Exactly as in the two-person Ultimatum Game, the prediction of the rational self-interest hypothesis in this game with competition among proposers is that all proposers will offer a 100 percent share, which is accepted by the responder. Hence all proposers get a payoff of 0 and the monopolistic responder captures the entire surplus. This distribution of the surplus is lopsided but describes precisely what happened in the experiments. After a few periods of adaptation, the offered share was very close to 100 percent, so that all the surplus was captured by the responder.

This striking result does not reflect a difference between the preferences of these players and the preferences of players in the Ultimatum Game. Roth et al. also conducted the standard two-person Ultimatum Game in these four cities. In all cities the proposers proposed an amount much nearer 50 percent of the surplus, and in all countries low offers were rejected at a substantially higher rate than higher offers. The reason for the surprising result is not preferences but the competitive setting. In this strategic setting, preferences for fairness cannot have any effect.¹ To see this, suppose that each of the proposers strongly dislikes to getting less than the responder. Suppose also that he believes the highest offer by others will be s^* . Then if he offer less than s^* , his offer has no effect and he will get a monetary payoff of 0 with certainty. Further, he cannot

¹ My summary of the proof follows Fehr and Schmidt (2003).

prevent that the responder gets s^* and that one of the other proposers gets $1-s^*$; so he will suffer from getting less than these two. However, if he offers a bit more than s^* , then he will win the competition, get a positive monetary payoff, and reduce the inequality between himself and the responder. Hence, he should try to outbid his competitors. This process drives the share that is offered by the proposers up to 100 percent. There is nothing the proposers can do about it, even if all of the them have a strong preference for fairness.

Competition is thus a key element that affects the impact of fairness. But competition alone is insufficient to eliminate the impact of fairness. Fairness matters a great deal in games in which not all aspects of the exchange can be completely determined by contracting, or in which enforcement is very costly. Thus, the amount of tax evasion is likely to be affected by the perceived fairness of the tax system,² and compliance with the law more generally is affected by the perceived fairness of the process (Lind and Tyler 1988, Fehr Gächter and Kirchsteiger 1997). Moreover, in nearly all contracts with labor, information is imperfect. In a labor contract, it is difficult to monitor the level of diligence and effort provided by the employee. In such settings, the effect on the work morale of employees of the perceived fairness of the firm's policy may make direct cuts in wages unprofitable (for a discussion of experimental evidence, see Fehr et al 1998, and Kahneman, Knetsch, and Thaler 1986). Box 1 describes a case study of one firm that provides evidence of this effect. As the authors of this study note, however, "we would recommend that the reader exercise caution in extending our results to other settings; our paper provides a detailed case study of only one firm in one unique

² Frey and Weck-Hanneman 1984, Alm, Sanchez and de Juan 1995, and Andreoni, Erard and Feinstein 1998.

period of its history” (Krueger and Mas, 2004, p. 257).

Box 1. Labor strife and the production of defective Bridgestone/Firestone tires

Since the 1940s, Firestone tire plants had adhered to the industry-wide pattern bargain with labor unions. However, when negotiations for a new contract began in 1994, Firestone/Bridgestone proposed deviating from the industry-wide bargain by moving from an eight- to a 12-hour shift that would rotate between days and nights and by cutting pay for new hires by 30 percent. This request for concessions came at a time of falling national unemployment and just after a profitable year for the employer. Almost immediately after the union at the Decatur, Illinois plant called a strike, the company hired replacement workers. Labor strife at the Decatur plant closely coincided with lower product quality. Defective tires were related to 271 fatalities and more than 800 injuries until the recall of Firestone tires in 2000.

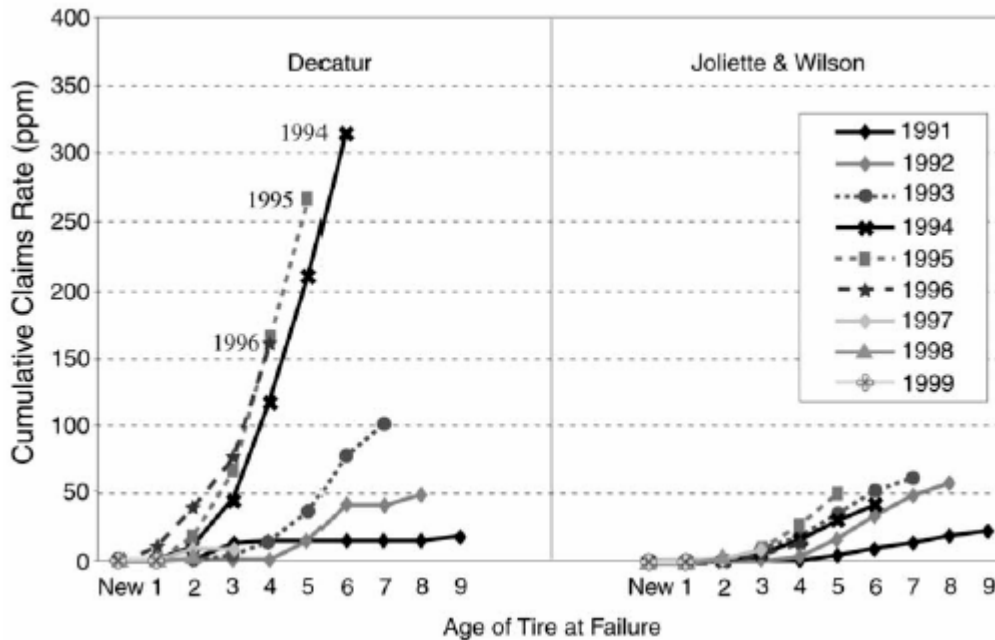
Tires are still made in large part by hand, so there is scope for human error in producing this product. Was the cause of the high rate of defective tires faulty workmanship by novice replacement workers? Was it a change in the technology of producing tires? Or did the labor strife play a role, with workers providing less effort and diligence because they felt they were not treated fairly?

In the 1990s, almost all Firestone ATX tires were produced in three plants: Decatur, Illinois; Joliet, Quebec; and Wilson, North Carolina. In Decatur, for nearly three years—from April 1994-December 1996—tires were produced by replacement workers, union members who crossed the picket line, management, and recalled strikers.

The Wilson plant was nonunion and so did not experience a strike. A Canadian union represented the Joliette plant, but labor relations there were much less contentious.

Joliette had a six-month strike over fringe benefits at the end of 1994, but the plant did not hire replacement workers (which are illegal in Quebec).

Figure 2: Complaints registered by the NHTSA per million P235/75R15 Firestone tires produced, by age of tire, production year, and plant (average of 10 imputed data sets).

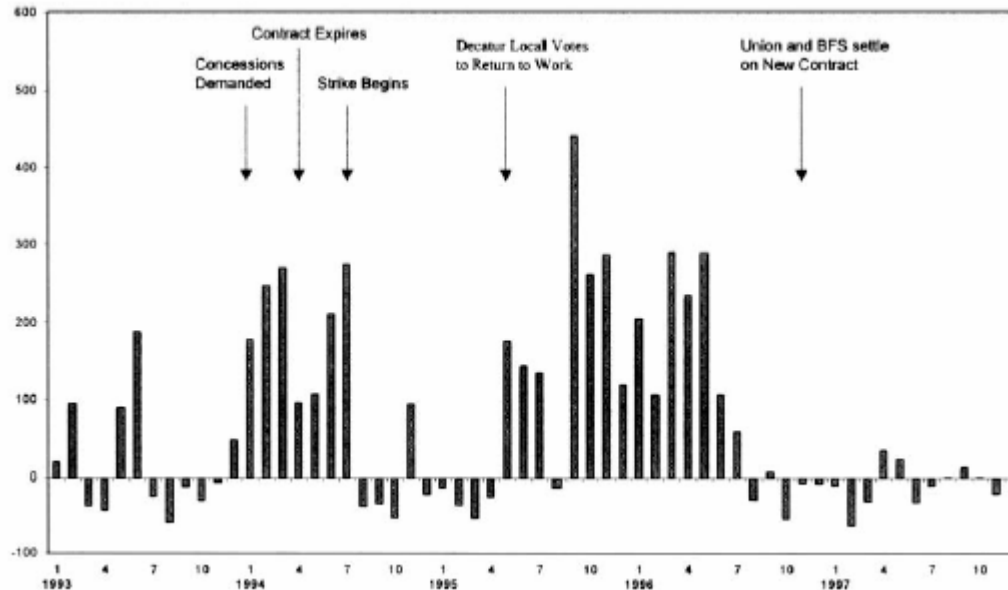


Source: Authors' calculations based on NHTSA and Safety Forum data. Complaints concern defects in 1991-99, reported prior to August 2000.

Figure 2 reports the cumulative number of complaints against Bridgestone/Firestone, per million ATX tires produced, by the year the tires were produced and the age of the tires. The left panel displays data for the Decatur plant and the right panel for Joliette and Wilson, the two other main producers. Tires produced during the labor dispute (1994 – 96) at the Decatur plant have a much higher failure rate than those produced at Joliette or Wilson, although before and after the dispute period the rate of complaints is similar for tires manufactured in Decatur and in the other plants. The three years of the labor dispute stand out in the figure for Decatur. The pattern suggests

that changes in technology cannot explain the rise in complaints against Decatur tires, as no such rise occurred in Joliet or Wilson.

Figure 3: Difference in the number of complaints per million tires produced, by month (Decatur plant minus Joliet and Wilson plants)



Source: Authors' calculations based on NHTSA complaints data. Records with missing data are excluded.

To try to isolate the effect of labor strife, it is useful to look at monthly data. Figure 3 displays the difference in the complaints rate by month of production between the Decatur plant and Joliet and Wilson plants. There was a spike in the first half of 1994, around the time in which concessions were demanded and the old contract expired. Through early 1995, when a large number of replacement workers were building tires, there were no excess complaint rate in the Decatur plant. It was not until the end of 1995, when large numbers of replacement workers and returning strikers worked side by side, that the excess number of complaints became high. These results suggest that faulty workmanship by novice replacement workers is not the sole source of problem tires. It appears likely that the chemistry between the replacement workers and recalled strikers,

or the cumulative impact of labor strife in general, created the conditions that led to the production of many defective tires.

Source: Krueger and Mas (*Journal of Political Economy* 2004 (2))

2. Punishing Cheats

A person's emotions may commit him to act in certain ways, whether or not it is in his self-interest to do so. As Robert Frank (1988) emphasized in his popular book, *Passions within Reason: The Strategic Role of the Emotions*, this commitment to deviate from self-interest can actually permit individuals to *advance* their self-interest. A person's emotional commitments to punish unfairness can in certain settings confer genuine material advantages because emotional commitments act as a commitment device that resolves social dilemmas--for example, problems of collective action or voluntarily supplied public goods. Social norms of behavior are an expression of emotional commitments that individuals should act in a certain way. A social norm is a behavior whose violation elicits some kind of social action (ostracism, unpleasant gossip, etc.). When people have a strong aversion to being treated unfairly, cooperation norms are likely to exist and individuals will be willing to punish free riding even if this is costly and they cannot expect future benefits from doing so. The willingness of individuals to bear a small cost to punish violations can have an important effect on behavior, since small punishments aggregated across many punishers can create a large cost for violations of norms.

A simple setting that captures some elements of the situation is the public goods game. In a public goods game, each of a fixed number of players chooses how much of his own resources to contribute to a public good that is shared by everyone. There is a

high group benefit to contributions to the public good, but for an individual player, not contributing (and free riding on others' contributions) is better. Fehr and Gächter (2000) conduct a public goods experiment under two conditions--with and without opportunities to punish free-riders. By using the no-punishment treatment as a control for the punishment treatment, it is possible to explore the power of individuals' aversion against being the "sucker"—that is, cooperating while others free ride.

In the game, each of four players receives 20 "monetary units" (say, US dollars) and can invest any portion of them in a public good shared by everyone. For every dollar that he invests, a player receives a benefit of 40 cents and so does every other player. Thus, a \$1 investment yields the group a benefit of \$1.60, which is evenly divided among them.

To see this clearly, designate the four players by A, B, C, and D. If player A contributes g_A , and the other three players contribute g_B, g_C, g_D , then player A's payoff is $20 - g_A + .4[g_A + g_B + g_C + g_D]$. Since a player loses 60 cents for each dollar he contributes, no one will contribute in this setting if all players are rational and self-interested. Then each player will just get \$20. If however all contribute their full endowment, then each player will get \$32. If one individual contributes and no other person does, he receives the "sucker" payoff of 0 and all others receive \$28. We summarize these results in Table 1.

Table 1. Payoffs in the Public Goods Game without Punishment

Contribution to the public good	Payoffs to:	
	Contributors	Non-contributors
0 by each player	-	\$20
Player A contributes \$20, other players contribute nothing	0	\$28
All contribute \$10	\$26	-
All contribute \$20	\$32	-

Fehr and Gächter conduct this experiment in the following way. Subjects interact in the laboratory and are told that they will play the game for 10 rounds; in each round, each player will interact with a different set of three players. After each round, subjects are informed of the total contribution of all players. This is the “no-punishment treatment.”

After the tenth round, subjects are informed that there will be a new experiment and that this experiment will again last exactly for 10 rounds. In the new experiment, subjects are given the opportunity to simultaneously punish each other after they are informed about the individual contributions of the other three members. A group member can punish another group member by assigning “punishment points” to him. For each punishment point assigned to an individual, his payoff is reduced by 10 percent. The cost to the punisher is \$1 for the first point, \$1 for the second point, and for subsequent points the cost rises slightly.

Figure 4a: Average contributions over time in the Stranger-treatment (Sessions 1 and 2)

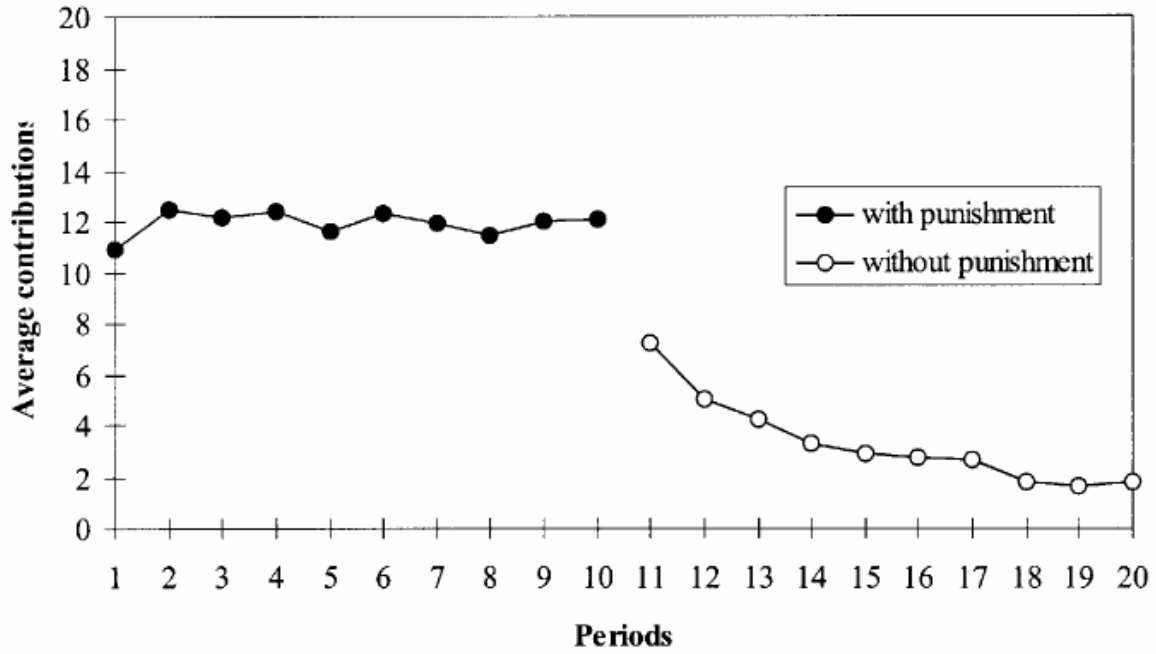
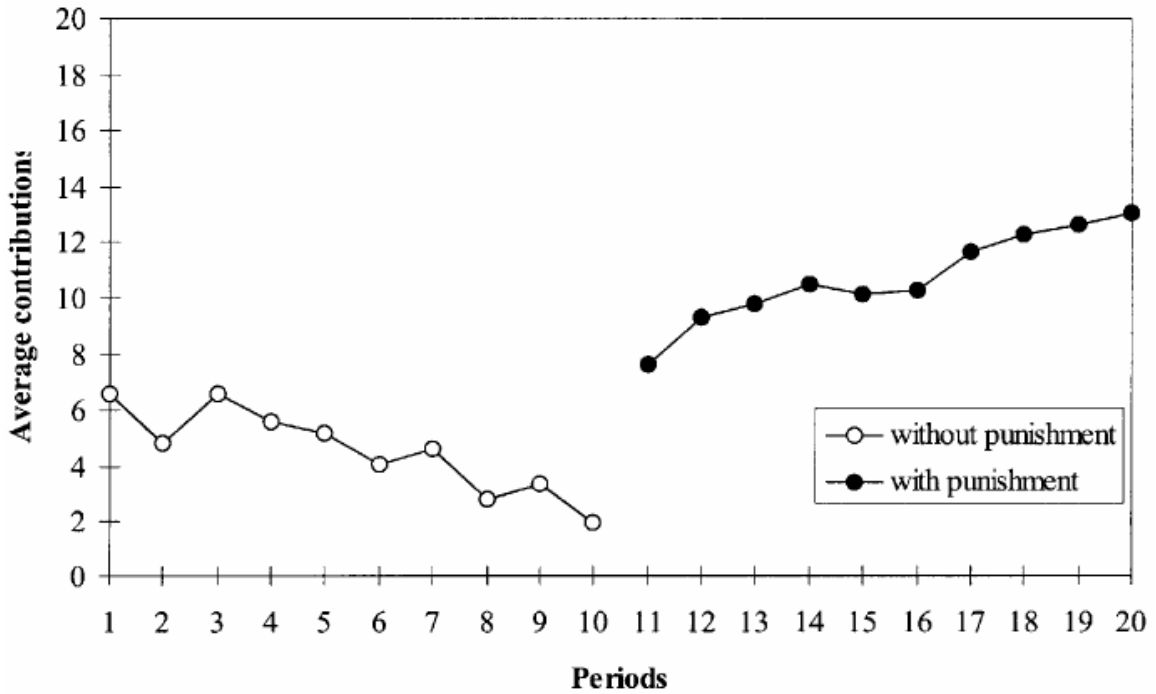


Figure 5b: Average contributions over time in the Stranger-treatment (Session 3)



Figures 4a and 4b report the results. In the bottom panel, the 10 rounds of the game without punishment are followed by 10 rounds of the game with punishment. In

the top panel, the order was reversed. The white dots show the typical pattern of average contributions over time when there is no scope for punishment. During the first period, average contribution levels were more than one-third of the endowment. However, in the final period, many players chose to contribute zero, and the average contribution was 10 percent. The fact that at the beginning of the public goods game without punishment, one regularly observes significant cooperation suggests that not all people are driven by pure self-interest. This and other experiments suggest that some individuals are conditionally cooperative—they will cooperate to some extent if others do. But once they see others who do not, they are unwilling to cooperate. Non-cooperators will lead nearly all individuals not to contribute. Fehr and Gächter (2000) conjecture that that is because subjects have a strong aversion to being the “sucker”—being those who cooperate while others free ride.

What is the effect of introducing punishment into the public goods game? Under the rational, self-interest hypothesis, no one would punish since interactions with a given player are not repeated and so he can have nothing to gain by sanctioning free-riding. Since a self-interested subject would not punish in this setting, creating scope for punishment should not affect behavior. However, the figure shows that the possibility of punishment causes a large rise in the average contribution in each round. This result holds irrespective of whether subjects play the punishment condition after or before the no-punishment condition; see the bottom and top panels of the figure, respectively. At the switch points between the treatments, there is a large gap in contributions in favor of the punishment condition. This suggests that some potential free riders anticipate that there are subjects who are willing to punish free riding. The results of this experiment

(not shown in the figure) also indicate that a subject is more heavily punished the more his contribution falls below the average contribution of other group members.

In this experiment, spontaneous punishment activities give rise to heavy punishment of free riders. This costly punishment occurs even though it provides no future benefit to the punisher. Potential free riders expect it. Punishing behavior causes a large rise in cooperation levels. With scope for punishment, a minority of altruists can force a majority of selfish individuals to cooperate. Without scope for punishment, a few egoists can induce conditional cooperators to defect.

A more general implication can be drawn. The scope for eliciting cooperative behavior may depend heavily on social norms—on eliciting the public’s role as “little policemen.” Many political scientists have asked the question: why do some governments act according to laws? Weingast (2003) provides the following answer. Political officials obey the law when citizens possess the ability to react in concert to potential transgressions by the state. “The fundamental problem of the rule of law is one of citizens policing the state.” When citizens come to expect a certain level of performance, and agree on what constitutes a transgression that violates a social norm, and when citizens are well-informed, then officials will develop a self-interest in adhering to the rules, which can make the rules self-enforcing.

A further implication is that if beliefs and norms can be changed, then such changes might play a role in bringing about effective institutional reform. The success story of a health care reform in northeast Brazil described in Tandler (1997) can be understood from this perspective. Tandler’s case study reveals a three-way dynamic among an activist state government, historically corrupt municipal governments, and

historically largely inert civil society. The state government conducted a program of mass advertising on the radio and in print that created in the community an expectation that the municipio was capable of delivering good health service. Civil society could then play the role of watchdog over the actions of local government and health care providers. This changed the environment in which the municipal governments functioned, introduced yardstick competition among them, and appears to have contributed to major improvements in public health in Ceara, one of the poorest states of Brazil.

The flip side of this argument is that when institutional reform is not accompanied by changes in social norms, the prospect for engineering successful institutional change may be limited. Successful institutional reform may require both formal changes that coordinate expectations, and informal changes in norms that lead individuals to punish deviations from the new formal standards.

3. Self-serving bias and bargaining impasse

All theories of fairness are based on the idea that actors have a criterion of fairness and compare themselves with a set of reference actors. In this section, we consider pitfalls that arise from the (unconscious) self-serving bias with which people adopt a criterion of fairness. In the following section, we consider the effect of historical inequalities on the characterization of who is like oneself, and who is not.

Standard bargaining models, which are based on the rational, self-interest hypothesis, attribute bargaining impasse solely to incomplete information. However, experimental evidence indicates that self-serving bias is also a factor in bargaining impasse. Babcock and Loewenstein (1997) conduct an experiment that pinpoints the

causal influence of self-serving bias. In the experiment, subjects bargained over how to settle a tort case involving a motorcycle accident. Their rewards depended on the settlement or, if they were unable to settle, on the actual court decision less a substantial cost (“court fees”). In the control treatment, individuals were assigned a role as plaintiff or defendant *before* reading the 27 pages of materials from the original legal case. In the experimental treatment, individuals were assigned a role *after* reading the legal case materials. The results were quite different in the two cases. When the subjects knew their role before reading the case materials, they were less likely to settle in face-to-face bargaining; they had formed quite different impressions from reading the case material compared to those who did not know their role in advance, and exhibited biased beliefs based on the position they were in. The greater the bias, the less likely they were to settle. The conflict in their beliefs was generated through nothing more than the roles they were assigned. Self-serving bias can be “self-defeating” bias, which blocks the achievement of mutually beneficial agreements.

This conclusion is borne out in a study of public sector contract negotiations. An important influence on bargaining positions in labor contract negotiations are agreements that have been made in “comparable” communities. The problem is that self-serving bias may enter into the each side’s choice of which communities are comparable. Babcock, Wang and Loewenstein (1996) surveyed union and school board presidents to obtain a list of districts that they viewed as comparable for purposes of salary negotiations. The districts listed by the two sides differed in a way that reflected a self-serving bias. The average salary in districts listed by the union was significantly higher than the average salary in districts listed by the board. In a regression analysis across school districts,

Babcock et al. found that a district where the average salary of the union's list is \$1000 greater than that of the board's list will be 49 percent more likely to strike than a district where the average salaries of the union's and board's lists are the same.

The greater the variance in the salaries of teachers in the neighboring districts, the greater the difference in the list of comparables was likely to be. However, the difference in the list of comparables was not related to the level of experience of the union or board present. In this study, experienced negotiators were as likely to succumb to self-serving bias as inexperienced ones.

A reliable finding in the economics literature is that social fragmentation hurts the delivery of public goods and, through that channel, economic performance.³ While further research is needed, the results of experiments on bargaining impasse may provide a micro-foundation for that result.

4. Durable Inequalities across Social Groups

Where do preferences over fairness? This question is at the frontier of economic research. One major joint research effort between anthropologists and economists put the question as follows (Hinrich et al. 2003):

“Are the violations of the selfishness axiom seen in experiments evidence of universal social preferences? Or, are social preferences modulated by economic, cultural and social environments? If the latter, which economic and social conditions are involved? ...Are there cultures that approximate the canonical account of purely self-regarding behavior?” (p. 1)

In Hinrich's study, the Ultimatum Game was conducted in fifteen small-scale societies. It was played for stakes that were the local equivalent of one or more day's wages. I summarize some of their findings below:

³For example, Easterly and Levine (1997), Alesina et al. (1999) and Banerjee and Somanathan (2001).

1. There is no society in which experimental behavior is consistent with the *rational, self-interest hypothesis*.
2. Differences among societies in market integration and the economic importance of cooperation explain a substantial proportion of the behavioral variation among groups. Societies with more market integration are more concerned about fairness, whereas in societies where economic integration is low, proposers in the ultimate game offered very little (the equivalent of \$1.50 out of \$10) and responders accepted virtually every offer.

In this section, we address a further question: If social preferences are modulated by economic, cultural and social conditions, then what is the effect of high historical levels of inequality on beliefs about fairness? Many historians, from Arnold Toynbee (1934) to Lauren Benton (2002), have noted that the dominance of one social group over another “uniformly sets in motion a process of cultural distancing...” (Benton, p. 13, cited in Hoff and Pandey 2004b). In Latin America,

“[t]he mere act of extending a claim of possession over American Indians ...changed Spaniards’ representation of Indians’ nature. The signs of civility and of shared humanity marveled at by Spaniards in their first encounters faded to insignificance after the formal act of possession...” (Benton, p. 12)

Evolving under the influence of economic and political institutions that created extreme inequality, belief systems in these societies represented the subordinate groups as *innately* inferior.⁴ Belief systems that are widely held change slowly. Belief systems

⁴ Many scholars have commented on why this was so. Sociologists have argued that in the absence of an ideology that represents the subordinate group as intrinsically inferior, an equilibrium of subordination cannot be maintained; see, for example, Pieris, 1952. Historians of colonialism, such as Dirks (2001) and Benton (2002), have argued that the structure of legal authority under colonialism was intertwined with the creation of a cultural hierarchy.

might be a vehicle for maintaining the persistence group inequality that we observe—for example, between blacks and whites in the United States, between former “untouchable” castes and all other castes in India, and between indigenous groups and settler groups in Latin America. Two dimensions of belief systems that appear to be important for the perpetuation of group inequalities are *social distance* and *stereotypes*.

(a) *Social distance*

[*This section is not yet written.*]

(b) *Stereotypes*. Psychologists have found evidence that unequal status breeds prejudice; individuals that because of social practices have unequal outcomes come to be viewed as innately unequal (for example, Hoffman and Hurst (1990)). Stereotypes affect the way people process information in ways that reinforce the stereotype.⁵ An experimental framework that reveals the effect of stereotypes on the way people process information is found in Stone, Perry, and Darley (1997). They asked all participants (who were American Caucasians, hereafter referred to as White) to listen to the same running account of an athlete’s basketball performance on the radio. Half the participants were led to believe that the target player was White, and half that he was African-American. The results indicated that information was less likely to be absorbed if it was discordant with the prevailing U.S. stereotypes that Whites are more academically talented than African Americans, and that African Americans are more athletically gifted. The White target player was perceived as exhibiting less natural athletic ability but more “court smarts,” whereas the African-American target player was perceived as exhibiting less court smarts but more natural athletic ability.

⁵ A survey of the literature on perceptual confirmation of stereotypes is in Hamilton and Sherman (1994).

Such biases have also been documented in practice. A recent study of the effect of stereotyping on judgment finds that prison inmates with more Afrocentric features receive harsher sentences than those with less Afrocentric features, controlling for race and criminal history (Blair, Judd, and Chapleau 2004).

A growing recognition of biases in perception has led economists to pay increasing attention to stereotypes as a mechanism behind the persistent gaps in economic achievement between observably distinguishable social groups long after overt barriers to economic and social mobility have been removed ⁶

Stereotypes can be self-fulfilling not only because they bias perceptions, but also because they influence the behavior of the stereotyped individuals. In another experiment, Stone et al. (1999) asked college undergraduate volunteers to complete a miniature golf course. The students' performance was measured by how many strokes were needed to put the ball in the hole. (Fewer strokes mean a better performance.) The variable that the experimenters manipulated was the description of the task. In one treatment, the task was described as a "standardized test of natural athletic ability." In the other treatment, the task was described as a "standardized test of sports intelligence." Table 2 presents the results. When the task was described as a test of natural athletic ability, the African-American participants performed better than the Whites: They averaged only 23.1 strokes to complete the 10-hole golf course, compared to an average of almost 27.8 for the White participants. But when the task was instead described as a test of intelligence, the race gap was reversed.

⁶See Loury (2002) for a wide-ranging study.

Table 2. Effects of Race and the Test Frame on Performance

	<u>Test frame manipulation</u>	
	Natural ability	Sports intelligence
Black participants, golf strokes	23.1	27.2
White participants, golf strokes	27.8	23.3

Source: Stone et al., 1999.

One way to interpret the behavior captured in the table is that social ideas—stereotypes about the talents of different social groups-- impose *bounds from within*. Whereas under the *rational, self interest hypothesis*, individuals change their behavior only when their preferences or external constraints change, the behavior of real individuals depends as well on belief systems that society has sedulously impressed on them. Negative stereotypes create anxiety that may interfere with performance; that is why the psychologist Claude Steele termed this kind of behavior “stereotype threat.” Positive stereotypes boost self-confidence that may lead individuals to expend greater effort.

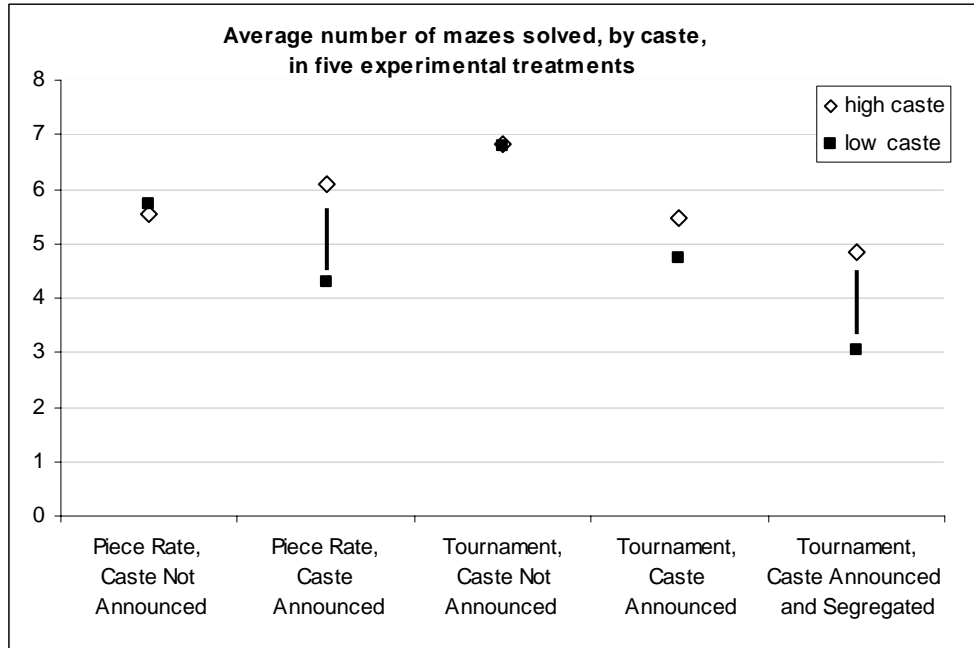
Stereotypes influence behavior twice—via their impact on individuals’ self-confidence, and also via their impact on the way individuals expect to be treated. To try to isolate the effect of making social identity salient on expectations of unfair treatment, Hoff and Pandey (2004a, b) undertook a series of experiments on Indian caste. The caste system in India can be described as a highly stratified social hierarchy in which largely

endogamous groups of individuals are invested with different social status and social meaning. A representative statement is Gupta (2000, p. 19, cited in Hoff and Pandey 2004a):

“Though there is no way by which those in a caste society can actually distinguish unfailing natural markers of difference, yet they justify caste stratification on the ground that different castes are built of different natural substances.”

In the first experiment, groups composed of three low-caste (“untouchable”) and three high-caste children were asked to solve mazes and were paid based on the number of mazes they solved. In the control condition, caste was not publicly announced. In a second condition, caste was made salient by publicly announcing each participant’s caste. In a third condition, individuals were segregated by caste and then their caste was announced in the six-person group. Figure 5 summarizes the results. There was no caste gap in performance when caste was not announced. However, increasing the salience of caste led to a significant decline in the average performance of the low caste, regardless of whether the payment scheme was piece rate or tournament. When caste was publicly announced, the low-caste children solved 23 percent fewer mazes on average in the piece rate treatments, compared to the performance of subjects when caste was not announced. In the tournaments, when caste was announced and groups were composed of six children drawn from *only* the low caste (a pattern of segregation that for the low caste implicitly evoked their traditional outcaste status), the decline in low-caste performance was even greater. A vertical line in the figure illustrates the statistically significant caste gaps.

Figure 6



Source. Hoff and Pandey (2004a)

In order to better understand the mechanism behind these results, in a new study Hoff and Pandey (2004b) manipulated the scope for discretion in rewarding performance. In this study, participants were asked to practice solving a puzzle based on the game *Rush Hour-Traffic Jam*. After practicing the game, they made a choice: They could accept a gamble in which they would attempt to solve a similar puzzle, with a payoff of 20 rupees (about US\$.45) if they succeeded and 1 rupee if they failed; alternatively, they could refuse the gamble and receive 10 rupees. In the control condition, the link between performance in the puzzle and the reward was mechanical;⁷ in the experimental condition, there was scope for discretion in rewarding performance because the frame

⁷ A player solves the puzzle by moving vehicles on a wooden game board in such a way that the “player’s car” escapes the gridlock and goes down the exit. The construction of the board makes it impossible to lift the vehicles off the board; they can only slide forward and backward on the tracks in the board. A frame on the game board prevents the vehicles from moving off the board except at the exit point. If a player solves the puzzle, his car exits and he obtains the success payoff, which is visible through the dashboard, from the underside of the car.

from the game board that kept vehicles on the road grid was removed, and a judge would award the success payoff only if the player did not let any vehicles move beyond the boundary of the game board. The subjects were told the following:

All around the roads are wheat fields. To solve the puzzle, you have to free the red car from the traffic jam by moving cars around in such a way that no car comes off the road. Otherwise the fields will be damaged... The person in the room who will give you the puzzle will watch you during that time. If he thinks you have solved the puzzle the right way, he will give you 20 rupees. If he thinks you have not solved the puzzle the right way, he will give you one rupee.

Consistent with a causal relationship running from caste salience to expectations of prejudicial treatment, the results showed that increasing the salience of caste had a large effect on the caste gap when there was scope for discretion in rewarding performance, but not when there was no such scope.

The expectation by the low caste subjects of prejudicial treatment may be rational given the discrimination that low-caste individuals experience in their villages. But the discrimination itself may not be fully rational. Cognitive limitations may prevent others from judging stigmatized individuals fairly.

The fact that people are bounded in their ability to process information creates broad scope for belief systems—in which some social groups are viewed as innately inferior to others—to influence economic behavior.

5. Conclusion

I briefly summarize the results.

- (a) Evidence abounds that individuals have preferences for being fairly treated and treating others fairly. These preferences do not affect economic outcomes in

- competitive markets with standardized products. They do affect economic outcomes in a wide variety of other settings where information is imperfect or enforcement is costly.
- (b) Emotional commitments to fairness norms can solve free rider problems. They may also play a role in maintaining a rule of law by enlisting citizens as informal policemen who make violations of the law costly for politicians.
 - (c) A pitfall of fairness preferences, combined with bounded rationality, is self-serving bias. Such biases can lead to costly bargaining impasse. In societies in the throes of massive social change, one may conjecture that there is no obvious single criterion of fairness, so that bargaining impasse may be more likely.
 - (d) Because many individuals value being treated fairly (relative to their reference set) and treating others fairly (relative to their social group), historical belief systems that create vast social distance between observationally distinct social groups, and that stereotype certain social groups as inferior, may tend to perpetuate inequality. The poorest members of a society are likely to be not only without economic resources, but also to be from stigmatized social groups, an identity that is likely to make their condition more difficult to change.

(e) References (preliminary; I would like to discuss all of these, but some have not yet been discussed in the current draft)

- Alesina, Alberto, Reza Gaqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114 (4): 1243-84.
- Babcock, Linda, and George Loewenstein. 1997. "Explaining Bargaining Impasse: The Role of Self-serving Biases." *Journal of Economic Perspectives* 11 (1): 109-26.
- Banerjee, Abhijit, and Lakshmi Iyer. 2003 "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." Mimeo, Massachusetts Institute of Technology, Department of Economics.
- Banerjee, Abhijit, and Rohini Somanathan. 2001. Caste, Community, and Collective Action: The Political Economy of Public Good Provision in India"Department of Economics. Massachusetts Institute of Technology, Cambridge, Mass.
- Blair, Irene V., Charles M. Judd, and Kristine M. Chappleau. 2004. "The Influence of Afrocentric Facial Features in Criminal Sentencing," *Psychological Science* 15(10): 674-679.
- Camerer, C. F. and R. H. Thaler. 1995 "Ultimatums, Dictators and Manners," *Journal of Economic Perspectives*, 9, 209-219
- Cameron, Lisa, 1999. "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia." *Economic Inquiry*, 37(1), 47-59.
- Camerer, F. Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Russell Sage Foundation, New York, New York. Princeton University Press, Princeton, New Jersey.
- Easterly, William and Ross Levine. 1997. "Africa's growth tragedy: Policies and ethnic divisions," *Quarterly Journal of Economics* 111 (4) 1203-1250.
- Fehr, Ernst and Simon. Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90, 980-994.

- Fehr, Ernst, G. Kirchsteiger, and A. Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics*, 108, 437-460.
- Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, 114, 817-868
- Fehr, Ernst and Klaus M. Schmidt (2003) "Theories of Fairness and Reciprocity— Evidence and Economic Applications," in M. Dewatripont, L.P. Hansen, and S. Turnovsky, eds. *Advances in Economics and Econometrics*. Vol. 1, Cambridge University Press, pp. 208-257.
- Fisman, Raymond and Shang-Jin Wei. 2004. "Tax Rates and Tax Evasion: Evidence from "Missing Imports" in China". *Journal of Political Economy* 112 (1): 471-496.
- Forsythe, R. L., J. Horowitz, N. E. Savin, and M. Sefton. 1994. "Fairness in Simple Bargaining Games," *Games and Economic Behavior*. 6. 347-369.
- Frank, H. Robert. 1988. *Passions Within Reason: The Strategic Role of the Emotions*. W.W Norton & Company, New York – London.
- Hamilton, David L., and Jeffrey W. Sherman. 1994. "Stereotypes." In R.S. Wyer, Jr., and T.K. Srull (eds.), *Handbook of Social Cognition*, 2nd Ed., Vol. 2, Hillsdale, NJ: Lawrence Erlbaum Associates, 1-68.
- Hinrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2003. *Culture and Social Preferences: Economic Experiments and Ethnography in 15 Small-Scale Societies*. Forthcoming book manuscript.
- Hoff, Karla, and Priyanka Pandey. 2004a. "Belief Systems and Durable Inequalities: An Experimental Investigation of Indian Caste." World Bank Research Paper 0-2875. Development Economics Research Group, Washington, D.C.
- Hoff, Karla and Priyanka Pandey. 2004b. "Belief Systems and Barriers of Mistrust: Experimental Evidence from India." Mimeo, World Bank, Development Economics Research Group, Washington, D.C.
- Hoffman, Curt, and Nancy Hurst (1990). "Gender Stereotypes: Perception or Rationalization?" *Journal of Personality and Social Psychology* 58 (2):

197-208.

- Hoffman, E. K. McCabe, and V. Smith. 1996. "On Expectations and Monetary Stakes in Ultimatum Games". *International Journal of Game Theory*, 25, 289-301.
- Kagel, J. H, C. Kim and D. Moser. 1996. "Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs," *Games and Economic Behavior*, 13, 100-110.
- Kahneman, D., J. L. Knetsch, and R. Thaler. 1986. "Fairness as a Constant on Profit Seeking: Entitlements in the Market," *American Economic Review*, 76, 728-741.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness and the Assumptions of Economics." In: *Rational choice: The contrast between economics and psychology*. Chicago and London: University of Chicago Press 1986, pp. 101-16.
- Krueger, B. Alan and Alexandre Mas. 2004. "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone / Firestone Tires". *Journal of Political Economy* 112 (2): 253-289.
- Ledyard, L. 1995. "Public Goods: A Survey of Experimental Research," Chapter 2, in *Handbook of Experimental Economics*, (ed. By A. Roth and J. Kagel), Princeton, NJ: Princeton University Press.
- Lind, A. and T. Tyler. 1988. *The Social Psychology of Procedural Justice*. New York: Plenum Press.
- Loury, Glenn C. 2002. *The Anatomy of Racial Inequality*. Cambridge, Mass.: Harvard University Press.
- Maravall, José Maria and Adam Przeworski. 2003. *Democracy and the Rule of Law*. Cambridge University Press, Cambridge, United Kingdom
- Ostrom, Elinor. 2000 . "Collective Action and the Evolution of Social Norms". *Journal of Economic Perspectives* 14, 137-158.
- Pandey, Priyanka. 2004. "Are Institutions Malleable? Effects of History, Mandated Representation and Democratization on Public Schools in North India." Working Paper. Pennsylvania State University and Indian Institute of Management, Lucknow, India.

- Rabin, Matthew. 2003. "Incorporating Fairness into Game Theory and Economics"
American Economic Review 83(5), pp. 1281-1302.
- Roth, A. E., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir. 1991. "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental Study". *American Economic Review*, 81, 1068-1095.
- Stone, Jeff, Christian I. Lynch, Mike Sjomeling, and John M. Darley. 1999. "Stereotype Threat Effects on Black and White Athletic Performance." *Journal of Personality and Social Psychology* 77 (6): 1213-27.
- Stone, Jeff, Zachary W. Perry, and John M. Darley. 1997. "White Men Can't Jump': Evidence for the Perceptual Confirmation of Racial Stereotypes Following a Basketball Game." *Basic and Applied Social Psychology* 19 (3): 291-306.
- Sunstein, Cass R. 1995. "Gender, Caste, and Law." In Martha Nussbaum and Jonathan Glover, eds., *Women, Culture, and Development*. Oxford, U.K.: Clarendon Press.
- Suleiman, R. 1996. "Expectations and Fairness in a Modified Ultimatum Game," *Journal of Economic Psychology* 17: 531-554.
- Tendler, Judith. 1997. *Good Government in the Tropics*. Baltimore: Johns Hopkins University Press.