The Carrot or the Stick: Rewards, Punishments and Cooperation*

James Andreoni
University of Wisconsin

William Harbaugh
University of Oregon and NBER

Lise Vesterlund
Iowa State University
and Harvard Business School

PRELIMINARY

Abstract

We examine the demands for rewards and punishments and their effects on cooperation. We consider a simple proposer-responder game where a proposer first makes an offer to split a fixed-sized pie, and then a responder is given the option of changing the proposer's payoff. Depending on the treatment the responder may at a cost to himself increase the proposer's payoff, decrease the proposer's payoff, or choose to either increase or decrease the proposer's payoff. We find substantial demands for punishments and rewards, and that the demand for rewards is not independent of treatment. Conditioning on the offer, rewards are larger when there is no option of punishing. We also find substantial treatment differences in the proposer's offers. Interestingly rewards do not work in the absence of punishments. While the proposer's offers are largest when the responder is able to both punish and reward, we find that the modal offer in the rewards-only treatment is the minimum offer.

*This paper was partly written while Harbaugh and Vesterlund were visiting the Center on Philanthropy and the Department of Economics at IUPUI. We are grateful for their hospitality and financial support. We also acknowledge support from the National Science Foundation. Finally, Vesterlund acknowledges support from ISU. We thank Yoella Bereby-Meyer, Gary Charness, Miguel Costa-Gomes, Rachel Croson, Muriel Niederle, Axel Ockenfels, and Al Roth for helpful comments and conversations.

1 Introduction

Economists have grown increasingly interested in studying individuals' concerns for others—sometimes called a preference for altruism or fairness. Understanding these preferences will help shape everything from abstract models of games to the design of institutions that take advantage of these preferences. Interest in these preferences is motivated in part by results from experiments, but also by experience in our daily lives. We all seem to care about the unwritten social contract that kindness should be shared and we are often willing to pay to enforce these ideals. For instance, every time we take a taxi or sit in a restaurant we entrust our happiness to another person. If that trust is protected then we may reward it with a generous tip, but if not we may punish with no tip at all. Likewise we may shun unfriendly colleagues but invite the friendly ones to our homes, and secretaries may perform more promptly for those who are polite or bring gifts.

Economic laboratory experiments have also found evidence of other-regarding preferences and willingness to both punish and reward. In static allocation games, such as dictator and public goods games, individuals show a clear willingness to share a surplus with others in a way that might be called altruistic. In sequential games, many individuals show a clear willingness to reward those who share and to punish those who have not shared enough.[1] Sometimes adding these rewards and punishments results in greater efficiency, but often not. In all these games, however, it is clear that individuals are extremely heterogeneous. Some individuals are perfectly selfish while others seem to have a strong aversion to inequality.

In this paper we begin a systematic look at both punishments and rewards and their effect on cooperation in economic laboratory experiments. We study a series of two-person proposer-responder games where subjects are continually rematched with new partners. We consider three treatments. In each treatment a proposer can share a surplus with the responder. In one treatment the responder can, at a cost to herself, either punish or reward the proposer. The next two treatments restrict the responder to strategies of reward-only or punishment-only, but are otherwise identical. This frame-

[1] To keep the exposition simple we will refer in two-stage games to a punishment as a decrease in payoff that a second-stage player imposes on a first-stage player. Similarly we refer to a reward as an increase in payoffs in the second stage. As examples of punishments see Güth, Schmittberger, and Schwarze (1982), Forsyth, Horowitz, Savin, and Sefton (1994), Bolton and Zwick (1995), Fehr and Gächter (2000), and Andreoni, Castillo and Petrie (2000). Examples of a demand for rewards are Berg, Dickhaut, and McCabe (1993), McKelvey and Palfrey (1992), Fehr, Kirchsteiger and Riedl (1993), Charness (1996), Charness and Haruvy (1999), Charness, Haruvy, and Sonsino (2000).

work allows us to identify the pure effects of rewards and punishments. By considering two-person games we avoid free-riding on punishments, and by randomly changing partners we avoid any repeated game effects. Moreover, by looking at rewards and punishments both individually and jointly we can identify any interaction or complementarity of the two tools. We hope to both identify the demands for rewards and punishments by responders and also to measure their impact on the cooperative behavior of proposers.[2]

We find that demands for both rewards and punishments are substantial. As expected, an increase in the offer by proposers, on average, decreases the punishment and increases the reward. Interestingly we find that while the demand for punishment appears to be independent of the reward option, the opposite does not hold for the demand for rewards. Conditioning on the offer, rewards are larger when the responder doesn't have an option of punishing. On the question of proposals, we find that the three treatments differ in their ability to generate generous offers. On average the proposed offer is largest when a combination of rewards and punishments are available and smallest when only punishments are available. Relatively speaking, however, the rewards-only treatment is the least effective in moving the proposers away from the minimum possible offer. This suggests that designing an institution around rewards only and forgetting an option for punishments may be a mistake.

In the next section we provide a theoretical framework for the study. We then present the experimental design in section 3 and the results in section 4. In section 5 we interpret our findings relative to explanations offered for fairness in experiments. Section 6 is the conclusion.

2 Background

A number of experiments examine the effect of either rewards or punishments. In this paper we focus exclusively on a limited set of these. Specifically, we only examine two player games where a proposer and responder

[2] To our knowledge we are the first to study these questions in a simple proposer-responder setting. Dickinson (1999) examines a team production problem where subjects may be exogenously rewarded or punished. Dickinson finds that this mechanism increases efficiency. Sefton, Shupp, and Walker (2000) examine repeated linear public goods games and find that the combined use of rewards and punishments result in the most generous public good contributions. Although not a formal study on the demands, Andreoni, Brown, and Vesterlund (1999) find results that suggest a tradeoff between the rewards and punishments. In a sequential public goods game they find that subjects punish small contributions but seldom choose to reward generous contributions.

must split a pie, Π, between them. The structure of the game is the following: the proposer first offers the responder a division (π_P, π_R) such that $a_R\pi_R + a_P\pi_P = \Pi$, where $a_R > 0$ and $a_P > 0$. Having observed the offer the responder may at a cost to himself punish or reward the proposer by choosing a vector (p, r), the costs per unit are respectively $c_p > 0$ and $c_r > 0$. The final payoffs of the game are $(w_P, w_R) = (\pi_P + r - p, \pi_R - c_p p - c_r r)$. Each game γ in this class can be characterized by the parameters $(a_P, a_R, \Pi, c_p, c_r)$. The standard subgame perfect equilibrium prediction for this type of game, independent of the parameters of the game, is simple: when it is common knowledge that all subjects aim at maximizing their own payoff, then the responder chooses $(p, r) = (0, 0)$, and the rational proposer offers $\pi_R = 0$.

The majority of the previously examined games in this category have been games that allow the responder only a single option of either punishing or rewarding, but not both.[3] When one of these options is unavailable we denote the cost of the excluded action as infinite. Excluding an option does not affect the standard equilibrium prediction: the responder does not punish or reward any offer, and the proposer keeps the whole pie.

Previous experimental results do not confirm this prediction. For example, in the frequently-studied \$10 dictator game, where $(a_P, a_R, \Pi, c_p, c_r) = (1, 1, 10, \infty, \infty)$, many have found an average offer of about 25%.[4] Deviations from the equilibrium prediction are also found in the ultimatum game. Here the responder is given an option of rejecting the proposer's offer and thereby decreasing the payoff of both to zero. Hence, if we restrict p to equal either 0 or π_P, the ultimatum game has the parameters $(a_P, a_R, \Pi, c_p, c_r) = (1, 1, 10, \frac{\pi_R}{\pi_P}, \infty)$. Figure 1 illustrates the set of possible payoffs in the game. The proposer's payoff is measured on the horizontal axis and the responder's payoff on the vertical axis. The proposer chooses an offer along the bold solid line, and conditional on that offer the responder faces a choice between the point chosen on the bold line and the origin. That is the responder may reduce the proposer's payoff at a cost to himself.

[3] One exception is Offerman (2000) who examines a proposer-responder game, where a proposer chooses between a hurtful or a helpful action, and the responder must select one of three possible payoffs: a cool response, a reward, or a punishment.

[4] See for example Forsythe, Horowitz, Savin, and Sefton (1994). Andreoni and Miller (1998) compare a series of games $(a_P, a_R, \Pi, \infty, \infty)$, which enables them to determine demands for $\pi_R = f(a_P, a_R, \Pi)$.

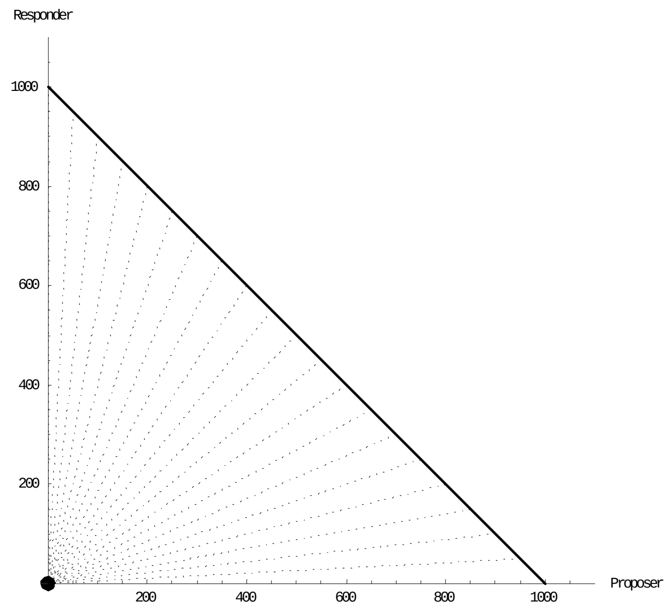

Figure 1: Possible Payoffs in the Ultimatum Game.

The average offer in the ultimatum game is generally around 40% and responders do indeed reject offers, however they are more likely to reject small offers than large offers. It is difficult, however, to determine the demand for punishments from these results. One reason is that the responder is only allowed one punishment: reducing the proposer's payoff to zero.[5] A second reason is that an increase in the proposed amount may not only increase the perceived generosity of the proposer, but it also increases the cost of rejection. If presented with a small offer it is relatively cheap for the responder to decrease the proposer's payoff substantially.

There have also been a series of experiments that examine how the possibility of a reward may affect the proposer's offer. The one game that clearly falls into the class of games that we are studying is the trust game.[6] In the trust game the proposer and the responder are each endowed with $10, and

[5] This restriction on the responder's choice set implies that the information revealed by his choice is quite limited. This may be one of the reasons why most of the current fairness models are consistent with the evidence from the ultimatum game (e.g. Bolton and Ockenfels (2000), Fehr and Schmidt (1999), Dufwenberg and Kirchsteiger (1998), Falk and Fischbacher (1998), and Rabin (1997).) See Andreoni et. al. (2000) for an anlysis of the convex ultimatum game.

[6] See Berg, Dickhaut, and McCabe (1995) for a careful description of this game.

the proposer is asked to make an offer to the responder. For every dollar transferred the responder receives three dollars, hence $a_P = 1$ and $a_R = \frac{1}{3}$. The responder may then choose to return any of the transfer to the proposer, therefore if we restrict $\pi_R \geq 10$, then the trust game γ^T is described by the following parameters: $(a_P, a_R, \Pi, c_p, c_r) = (1, 1/3, \frac{40}{3}, \infty, 1)$. Figure 2 illustrates the set of possible payoffs in this game. The bold line illustrates the proposer's choice set, and the thin lines show the possible constraints for the responder. The actual constraint will be the thin line that originates at the proposer's offer on the bold line.

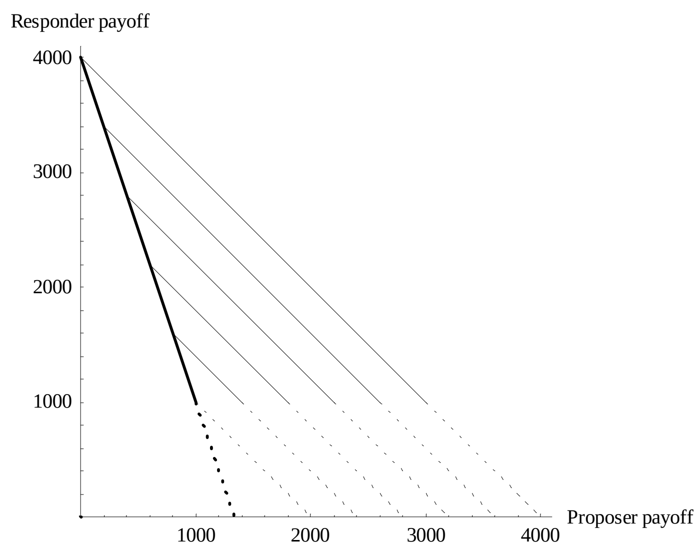

Figure 2: Possible Payoffs in the Trust Game.

Once again the experimental evidence does not support the standard equilibrium prediction. In the Berg, Dickhaut, and McCabe (1993) study, they find that the proposer passes an average of \$5.16, and responders return an average of \$4.66. Furthermore there is a positive correlation between the amount sent and the amount returned.

Overall the previous results show that responders are willing to punish and reward. The responses are typically correlated with the initial offer, π_R, and there is reason to believe that $p = f(\pi_R)$ and $r = g(\pi_R)$. In some of these studies, the offer itself changes other parameters of the game, and even when these are held constant it is generally the case that only one game is examined, implying that we are unable to determine what other factors may affect the demand for punishments and rewards.[7]

In this paper we report experimental results which help us identify some of the factors that affect demands for rewards and punishments. By studying rewards and punishments separately and jointly we are able to identify the possible interaction between the two. Our hope is that by better understanding what affects the demands, we will be able to gain insight to what motivates them.

In addition to determining the demands we also determine whether the responder's choice set affects the proposer's initial offer. One might expect that the combined availability of punishments and rewards is more successful in increasing the proposer's offer, but it is not clear what type of difference one should expect when comparing the options alone.

3 Experimental Design

We consider three variations of the general class of games described above. The only difference between them is in the availability of punishments and rewards. The specific parameters of the game are as follows. In the first stage a proposer decides how large a portion of \$2.40 he wants to transfer to the responder; where $a_R = a_P = 1$. Although we considered alternative costs we decided that for this initial study a couple of extreme prices would be useful. Using the notation from before, we consider costs of either ∞ or $\frac{1}{5}$. We refer to our three treatments as: *Carrot-Stick, Carrot,* and *Stick*. In the Carrot-Stick Treatment the responder can at a cost of one cent increase or decrease the proposer's earnings by 5 cents. In the Stick Treatment the responder can at a cost of one cent decrease the proposer's earnings by 5. Finally, in the Carrot Treatment the responder can at a cost of one cent increase the proposer's earnings by 5 cents. Hence the parameters of the Carrot-Stick treatment are $(a_1, a_2, \Pi, c_p, c_r) = (1, 1, 2.40, 1/5, 1/5)$; Carrot: $(a_1, a_2, \Pi, c_p, c_r) = (1, 1, 2.40, \infty, 1/5)$; and Stick: $(a_1, a_2, \Pi, c_p, c_r) = (1, 1, 2.40, 1/5, \infty)$.

The Carrot-Stick treatment is illustrated in Figure 3 below. The proposer chooses an offer along the bold solid line, and conditional on that offer

[7] Another example where the cost of punishing is dependent on the level of cooperation is Fehr and Gächter (1998). In their examination of the public goods game, subsequent punishments are a percentage of the individual's payoff, and as a result it is cheaper to punish those who are less generous.

the responder has the option of choosing any point on the reward and punishment line originating at the proposer's offer. These choices are indicated by the lighter lines. To secure the responder the opportunity of decreasing the proposer's payoff to zero, we do not allow the proposer to make offers below 40 cents.

The light dashed line indicates all possible payoff combinations that result in equal payoffs to the responder and proposer. Clearly the responder decreases her payoff by choosing any outcome off of the original proposal on the bold line, however note that relative to the bold line outcomes, areas I and III result in a more equal distribution of payoffs. We refer to subjects who choose outcomes in these two areas as equalizers. In contrast, subjects who pick outcomes in areas II and IV could have chosen an outcome which both resulted in a larger personal payoff as well as a more equal distribution of payoffs.

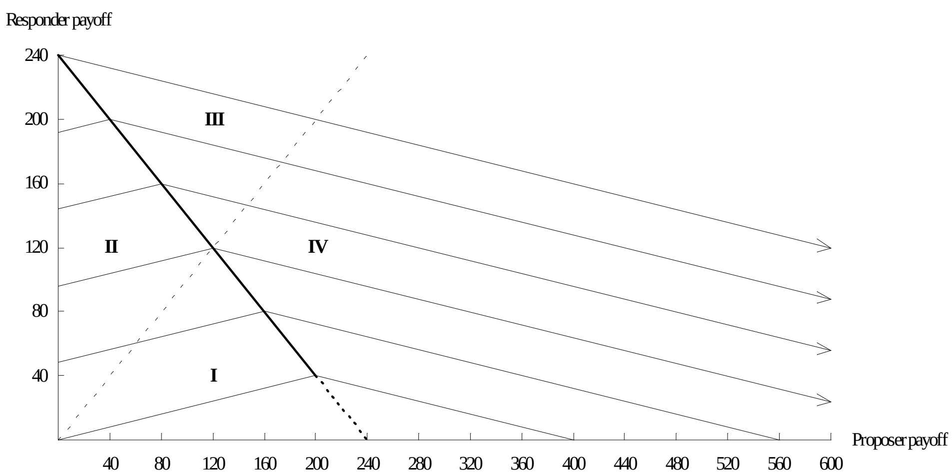

Figure 3: Possible Payoffs in Carrot-Stick

The Carrot Treatment allows for all payoff combinations on or to the right off the bold line, while the Stick Treatment allows for payoff combinations on or to the left off the bold line.

If it is common knowledge that the objective of all individuals is to maximize their personal payoff, then the subgame perfect equilibrium prediction is the same for all three treatments. The responder should neither reward nor punish, and given this response the proposer chooses the minimum required transfer of $0.40.

For each session we needed a total of 20 subjects. We therefore recruited a larger pool of undergraduate business students and selected the 20 sbuects at random from the pool. Those not selected for the experiment were paid $5 to leave. Subjects were randomly assigned to a computer terminal and were given a set of written instructions. The experimenter read the instructions aloud, following which the subjects were asked to calculate the payoffs in a specific example of the game. The answers to the quiz were collected, and the example was reviewed verbally by the experimenter. Half of the subjects were then randomly assigned to be proposers and half to be responders. The subjects were informed that they would remain a responder or proposer throughout the experiment. Subjects then began the experiment. They played 10 iterations of the game. In each iteration they were randomly and anonymously paired with another subject, with the stipulation that no one played another subject more than once. Subjects' identities were never revealed to one another. After the 10 rounds, subjects' earnings for all 10 rounds were tallied and added to a $5 show-up payment. While waiting for their payment, subjects were asked to answer a questionnaire. Subjects were paid anonymously with cash in "payment envelopes" which were handed out by subject number. We ran three sessions of each of the three games, for a total of 180 subjects. Per treatment we have 30 subjects in each role. The experiment typically lasted less than an hour, and including the show-up fee subjects made an average of $17.55 (standard deviation of 4.85, maximum of $49.35, and minimum of $6.70).[8] A copy of the instructions for the Carrot-Stick game can be found in the Appendix. We attempted to keep the instructions as neutral as possible and chose to refer to the punishments and rewards simply as changes to the proposer's payoff.

4 Results

In this section we first give a brief overview of the results and examine whether the observed behavior is consistent with the standard subgame perfect equilibrium prediction. This analysis reveals substantial differences across treatments. Next, we study these differences more carefully by examining the demands for rewards and punishments, and then by determining the extent to which these affect the cooperative behavior of the proposer. Finally we look at the distributions of final payoffs.

[8] The subject with the highest earnings was a proposer in the Carrot treatment. That subject made generous offers and received rewards of $36 over the course of the experiment. The subject with the lowest earnings was a proposer in the Stick treatment. This subject never made an offer below 200, and indicated in the questionnaire that her main objective was to avoid punishments.

4.1 Equilibrium Predictions

As noted earlier the standard subgame perfect equilibrium prediction is that there should be no punishments or rewards and that the proposer's offer to the responder should be 40 cents. Without surprise our data are not consistent with this prediction.

Let us first examine the proposer's behavior. The average offers in each of the ten rounds and for the three treatments are shown in Figure 4. In each of the three treatments and in each of the ten rounds we reject the hypothesis that the average offer equals 40 cents.[9]

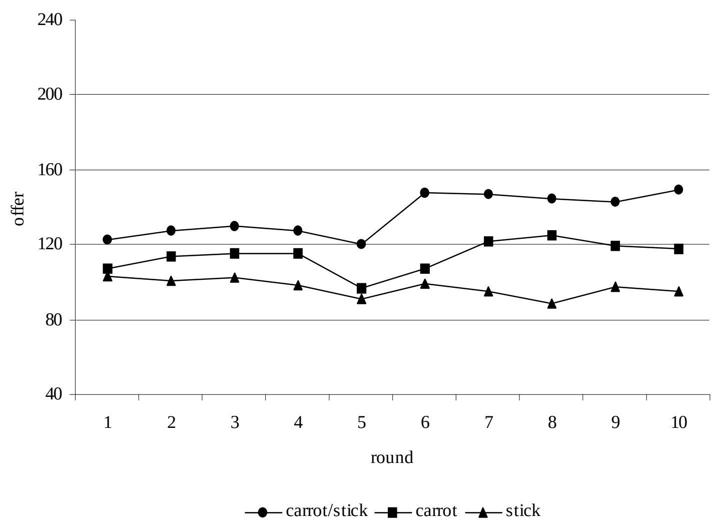

Figure 4: Average Offer.

As expected the offers do not decrease towards the end of the experiment, indicating that the proposers correctly understood the one-shot nature of the interaction.

Contrary to the equilibrium prediction that the offer is independent of treatment, the average offer is largest in the Carrot-Stick treatment and smallest in the Stick treatment. These differences are statistically significant and appear to be increasing over the course of the experiment.[10]

[9] Treating each proposer as an observation we determine whether the average proposal equals 40. Across the ten rounds the t-statistic is 12.6 for the Carrot treatment, 10.7 for the Carrot-Stick treatment, and 9.6 for the Stick treatment. The results for the last round and the last five rounds are similar.

[10] Treating each proposer as an observation we determine whether the average offers

Next we turn to the responder's decisions. Here we find little support for the hypothesis that there should be no punishments or rewards. We test whether the responder's cost of changing the proposer's payoff on average equals zero, and for all three treatments we reject the hypothesis.[11]

In fact the willingness to reward and punish is quite substantial. Across the three treatments responders chose to change the proposer's payoff 43 percent of the time, and over the ten rounds 80 percent of the responders chose at least one change of the proposer's payoff. During the last five rounds, 75 percent of the responders chose to change the proposer's payoff.

Figure 5 shows the proportion of offers that were either increased or decreased by the responder in a given period, and Figure 6 shows the responder's cost of changing the proposer's payoff.[12] In contrast to the repeated game result of Sefton et al. (2000) neither the likelihood nor the expenditure on changes decreases over time. This suggests that both proposers and responders behaved in a manner consistent with the one-shot interaction.

differ across treatments. For the ten rounds the test statistics are $t = 2.01$ for the Carrot and Carrot-Stick comparison, $t = 2.01$ for the Carrot and Stick comparison, and finally for the Carrot-Stick and Stick comparison, $t = 3.62$. Similar results are found for the last five periods. For the last period the difference between the Carrot-Stick and Stick treatments are significant at the five-percent level, whereas the others are significant at the ten-percent level.

[11] Treating each responder as an observation we determine whether responders on average chose not to change the proposer's payoff. Over the ten rounds the t-statistic is 4.6 for the Carrot treatments, 7.5 for the Carrot-Stick treatment, and 5.9 for the Stick treatment. The results for the last round and the last five rounds are similar.

[12] Given the random re-matching and a heterogeneous population it is not surprising to observe a substantial degree of fluctuation in these numbers. Suppose for example that there are two types of people, equal-dividers and free-riders. Free-riders never change the payoff and always offer 40 tokens, equal-dividers always offer 120, and as responders they choose a change that equalizes the payoff of the two participants. If free-riders are matched with free-riders and equal-dividers with equal-dividers, then the average desired change is zero. However if subjects are matched with their opposits, then the average change is 20 tokens.

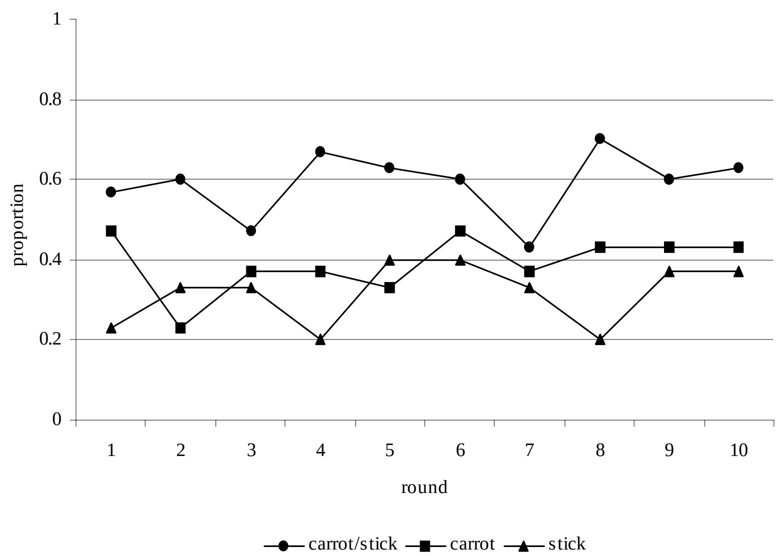

Figure 5: Proportion of Offers Changed.

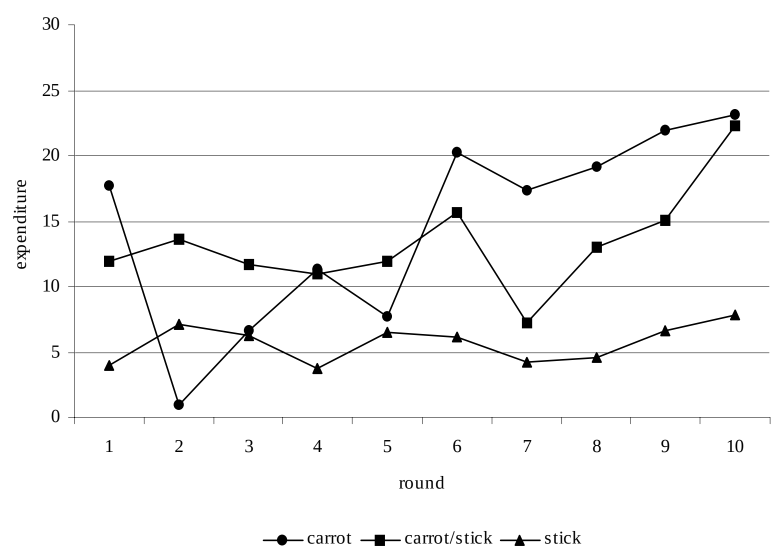

Figure 6: Average Expenditure on Changes.

Although less obvious there are some differences across treatments. During the last five periods the average responder in the Carrot treatment spends 20.4 cents changing the proposer's payoff. In comparison, the number in the

Carrot-Stick treatment is 14.7, and in the Stick treatment it is only 5.9.[13]

Given that neither the proposer nor the responder behaved according to the standard equilibrium prediction, it is of interest to ask what their payoff from this deviation was on average. Table 1 below shows that the proposer's payoff is largest in the Carrot treatment, while the responder's average final payoff is largest in the Carrot-Stick treatment.

Table 1: Average Final Earnings

	All Ten Rounds		Last Five Rounds	
Treatment	Proposer	Responder	Proposer	Responder
Carrot-Stick	126.4	122.5	122.5	131.5
Carrot	199.1	99.3	223.7	97.8
Stick	114.5	91.3	115.5	89.2

Relative to the standard equilibrium prediction, the chosen outcomes were both Pareto-damaging and Pareto-improving. The average joint payment per round is largest in the Carrot treatment where the proposer and responder jointly receive 298 tokens per round. In comparison the average joint payment is 249 in the Carrot-Stick treatment and only 206 in the Carrot treatment.[14] Of course there are also substantial differences in the relative payoffs to the proposer and responder. Relatively speaking the payoffs are more equally distributed in the Carrot-Stick treatment, and most unequal in the Carrot treatment.

Next we examine these treatment differences more closely. First, we examine if there is an interaction between demands for punishments and rewards. Following this analysis we examine whether and how these demands affect the proposer's offer. Finally we summarize our findings by looking at the actual distribution of final payoffs.

[13] Over all ten rounds the average cost of changing payoff are 14.7 in Carrot, 13.3 in Carrot-Stick, and 5.7 in the Stick treatment. Treating each responder as an observation we find that while the difference between the Carrot and Carrot-Stick treatment are not significant there is a significant difference between the Stick and the reward treatments. The t-statistic is 0.35 for the Carrot and Carrot-Stick comparison, 2.68 for the Carrot and Stick comparison, and finally, for the Carrot-Stick and Stick comparison it is 3.75. The results for the last period and the last five periods are similar.

[14] Determining whether the sum of payoffs on average are the the same across treatments we find that these differences are significant. Treating each responder as an observation we get the following t-statictics over the ten rounds: 3.02 for the Carrot and Carrot-Stick comparison, 6.62 for the Carrot and Stick comparison, and finally 3.60 for the Carrot-Stick and Stick comparison. Similar results are found for the last five rounds of the experiment, where the joint payoff in the Carrot treatment is 321, and only 254 and 205 in the Carrot-Stick and Stick treatments respectively. In the last period the difference between the Carrot-Stick and Carrot treatment is not significant.

4.2 Demands for Punishments and Rewards

The demands for punishments and rewards are substantial. More than half of the offers in excess of 120 result in an increase in the proposer's payoff. For offers below 120 more than 40 percent result in a decrease in the proposer's payoff.[15]

It is of interest to determine both how the demands for punishments change with the size of the proposer's offer, and also how these compare across treatments. To facilitate this comparison we truncate all the rewards in the Carrot-Stick treatment to equal zero. The demands for punishments during the last five periods are shown in Figure 7.

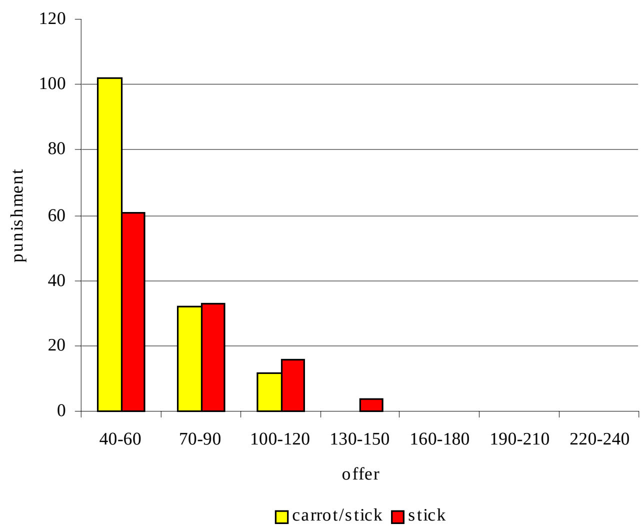

Figure 7: Average Punishment (last 5 rounds).

First we note that the average punishment decreases with the size of the proposer's offer. Hence we replicate the results that have been found when the cost of punishing is not independent of the offer. Second, the decrease in punishment is rather steep, and after the equal split there are

[15] Across all ten periods 54 percent of offers above 120 are rewarded in the Carrot treatment, and 59 percent of these offers are rewarded in the Carrot-Stick treatment. The proportion punished is somewhat lower. 43 percent of offers below 120 are decreased in the Stick treatment, and 51 percent of such offers are decreased in the Carrot-Stick treatment. Of the offers above 120, 8 percent are punished in the Stick, and 2 percent in the Carrot-Stick. Finally, of the offers below 120, 31 percent are rewarded in the Carrot, and 11 percent are rewarded in the Carrot-Stick.

essentially no punishments.[16] Of the 120-offers approximately 15% result in some decrease in the proposers payoff.[17] Finally, the responder's ability to reward has limited effect on the demand for punishments.[18] The punishments of medium-sized offer are the same across treatment, and although the punishments of low offers appear larger in the Carrot-Stick treatment, this difference is only significant at the end of the experiment.[19] This finding is surprising. One would perhaps not expect rewards of small offers, yet it is when the responder has the option of rewarding that the punishment of small offers is largest.

Next we examine the demand for rewards. Once again we truncate the data from the Carrot-Stick treatment by setting all punishments equal to zero. The average increase in proposer's payoff for the last five periods are shown in Figure 8. Similar to the evidence from the trust game a higher offer on average leads to a larger reward. In addition the data reveal an interesting result which one may not have expected based on the data from the trust game: average rewards at or below equal-split offers are quite substantial.[20] Another puzzling finding is the substantial differences between the Carrot and Carrot-Stick treatments.[21] Conditional on the offer

[16] Note that in the Stick treatment there are very few offers past the equal split.

[17] Of the 120-offers during the last five rounds 17% of them are decreased in the Stick treatment, and 13% are decreased in the Carrot-Stick treatment.

[18] To determine whether the responders' demands for punishments change when rewards are also available, we take a conservative approach to labeling observations as censored. We denote any observation that results in zero payoffs to the proposer, and any observation where the responder choose not to change the proposer's payoffs as censored. Using a random effects model we regress the truncated demand for punishments on the offer, a dummy for the stick treatment, and an interaction between the two, and account for censoring as described above. Looking across all ten rounds, or the last five rounds we cannot reject the hypothesis that the demands for punishments are the same across treatments. As expected an increase in the proposer's offer significantly decreases punishments.

[19] Counting each responder as an observation we test whether responders on average choose larger punihsments of low offers when the option of rewarding is available. Across the last five rounds the p-value of the test is 0.21, and for the last round the p-value is 0.10. It should be noted that in each of the last five rounds the punishment of low offers exceeds that observed in the absence of punishment.

[20] Such choices cannot be revealed in the trust game, where all transfers result in the responder receiving more than half the pie.

[21] Once again to determine whether the responders' demands for rewards change when punishments are available, we take a conservative approach to labeling observations as censored. Here we denote any observation that results in zero payoffs to the responder, and any observation where the responder choose not to change the proposer's payoffs as censored. Using a random effects model we regress the truncated demand for rewards on the offer, a dummy for the Carrot treatment, and an interaction term between the two, and account for censoring as described above. The joint hypothesis that there is no effect

the average reward given in the Carrot treatment is larger than the reward given in the Carrot-Stick treatment.[22] Even at very large offers where we know there is no demand for punishment, we see that the absence of the ability to punish results in a reward which is larger, on average.

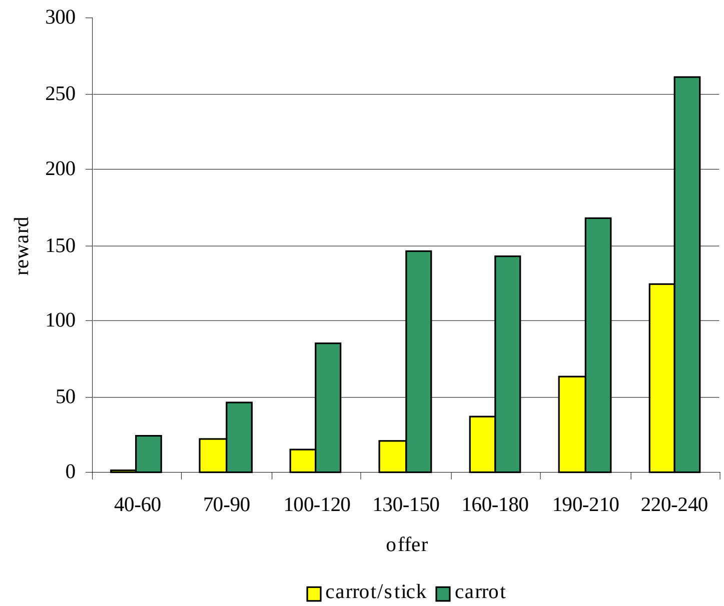

Figure 8: Average Rewards (last 5 rounds).

While punishments are unaffected by the availability of rewards, we find that the rewards are larger when there is no option of punishing. Note that it is the combination of these results that makes them puzzling. If one were presented simply with the reward results, then it might be tempting to argue that the difference is caused by a simple substitution effect. Suppose that rewards and punishments are viewed as substitutes, then it may be that the demand for either one is decreasing in own cost and increasing in the cost of the other input, i.e. $\frac{\partial p}{\partial c_p} < 0$ and $\frac{\partial p}{\partial c_r} > 0$, and similarly $\frac{\partial r}{\partial c_p} > 0$ and

from punishment option is rejected with a p-value of 0.04 over the ten rounds, and 0.03 over the last five rounds. As expected an increase in the proposer's offer significantly increases rewards.

[22] Treating each responder as an observation we test whether the average reward is larger in the absence of punishment option. The p-values for the last five rounds and for each of the seven offer ranges are from smallest offer to largest: 0.04, 0.30, 0.07, 0.03, 0.12, 0.17, and 0.05. The insignificant results are generally found for offers, where there generally are fewer observations.

$\frac{\partial r}{\partial c_r} < 0$. If a particular offer is perceived in a similar way across treatments, then excluding one of these substitutable tools simply suggests that rewards should be larger in the Carrot than in the Carrot-Stick treatment, $r_C > r_{CS}$. That is, if subjects are limited to only using one of two tools, then they will use the available tool more. In contrast to our experimental evidence this also implies that the demands for punishments will be larger in the Stick than in the Carrot-Stick treatment, $p_S > p_{CS}$. Therefore a simple substitution argument cannot explain the difference.

To summarize, holding cost constant we replicate the result that an increase in the proposer's transfer on average decreases the punishment and increases the reward. In addition we identify three puzzling phenomena. First, the rewards of small offers is substantial; second, rewards are larger when there is no punishment option; and third, the the reward option has little effect on the demand for punishments. In section 5 we determine the extent to which recent fairness models may help us better to interpret these findings.

Next we examine if the responders' choice set affects the proposer's offer to the responder.

4.3 Cooperation Production Function

Earlier we showed that the proposer's offer to the respondent differed significantly across treatment. During the last five rounds the average offer to the responder was 146 in the Carrot-Stick treatment, 118 in the Carrot, and 95 in the Stick treatment.[23] These numbers may be seen as evidence that the Stick treatment is least effective in moving the proposer away from the minimum required offer of 40 cents. However, when examining the actual distribution of offers we find that these simple averages are misleading. In fact the minimum required offer is more frequently observed in the Carrot treatment.

Figure 9 shows the distribution of offers over the last five rounds of the experiment. The modes of the distributions differ substantially across treatments: the most frequent offer in the Stick treatment is 120, whereas it is 240 and 40 in the Carrot-Stick and Carrot treatments, respectively.[24] Rewards alone are ineffective in moving the modal offer away from the minimum offer, whereas punishments are effective in moving the modal offer to an equal split offer. There is evidence that rewards and punishments may

[23] The average offers to the responder over all ten rounds are: 136, 114, and 97 respectively.

[24] The data for all 10 rounds is very similar to that seen for the last five rounds.

complement one another. While rewards alone are ineffective they are quite effective when combined with punishments. In the combined treatment we find that rewards are effective in moving the modal offer past the equal split.

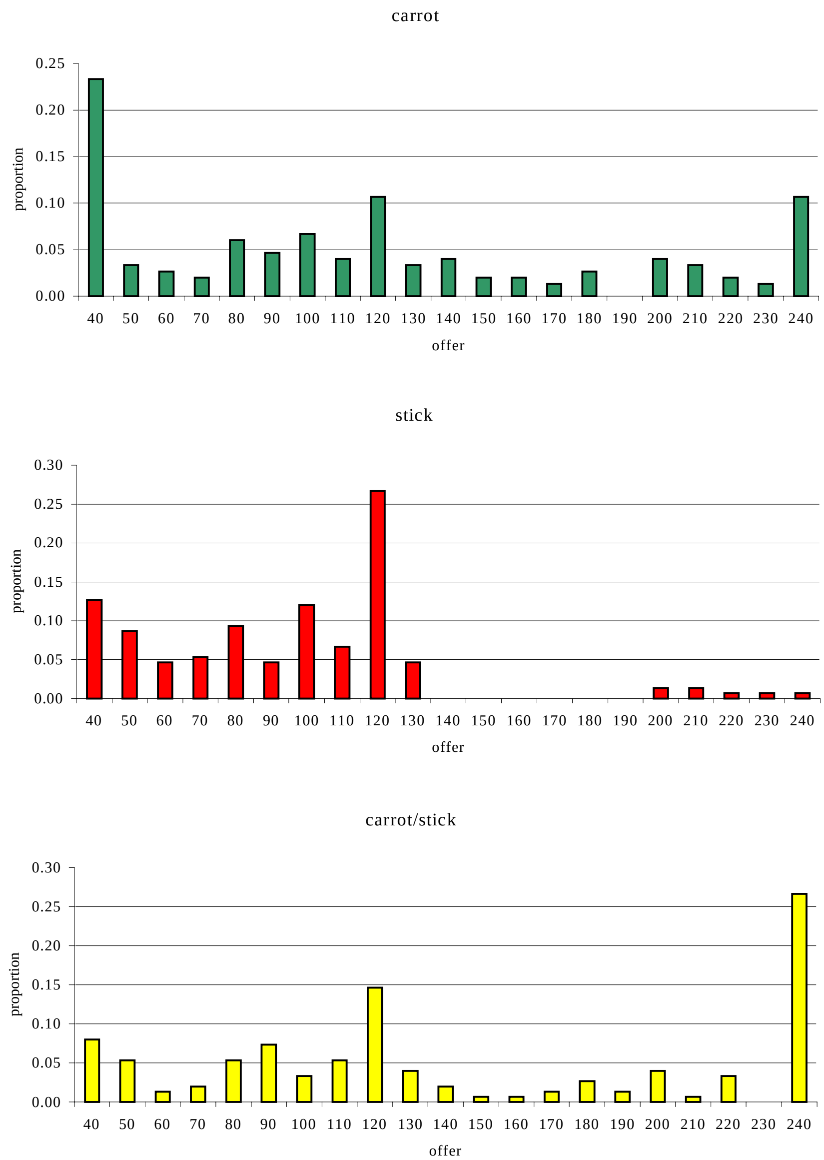

Figure 9: Distribution of Offers, π_R (last 5 rounds).

If the Carrot treatment is least effective in moving offers away from the classical subgame perfect equilibrium, then why is the average offer in this treatment higher than in the Stick treatment? The reason is that while a third of the offers in the Carrot treatment are past 120, practically no offers are seen in this range in the Stick treatment. This is perhaps not surprising given that punishments decrease dramatically as the equal split is reached. In fact the evidence from the Stick treatment is quite similar to that from the ultimatum game. The average and median offers are around 40 percent of the pie, and the mode of the distribution is about 50 percent of the pie.

4.4 Distributions of Outcomes

Finally, to summarize the results of our experiments we compare the resulting payoff distributions. We focus solely on the last five periods of the game.

Figure 10 shows the resulting outcomes from the Carrot-Stick treatment. The dark line illustrates the set of possible offers that the proposer could have chosen, and the thin lines originating at the relevant offer illustrate the possible changes that the responders could have chosen. The circles illustrate the observations, and the area of the circle, the number of observations. Circles on the right edge of the figure indicate observations where the proposer's payoff exceed \$4.00.

In the Carrot-Stick treatment forty percent of responders' choices do not change the proposer's payoff, however only 10 percent of the subjects prefer this outcome, independent of the offer. We also observe a substantial number of subjects choosing outcomes that simultaneously decrease equality and the responder's personal payoff - areas II and IV of Figure 3. A total of 23 percent of the observations during the last five rounds are in one of these two areas, corresponding to more than half of our subjects choosing at least one outcome which decreases their absolute and relative payoffs.

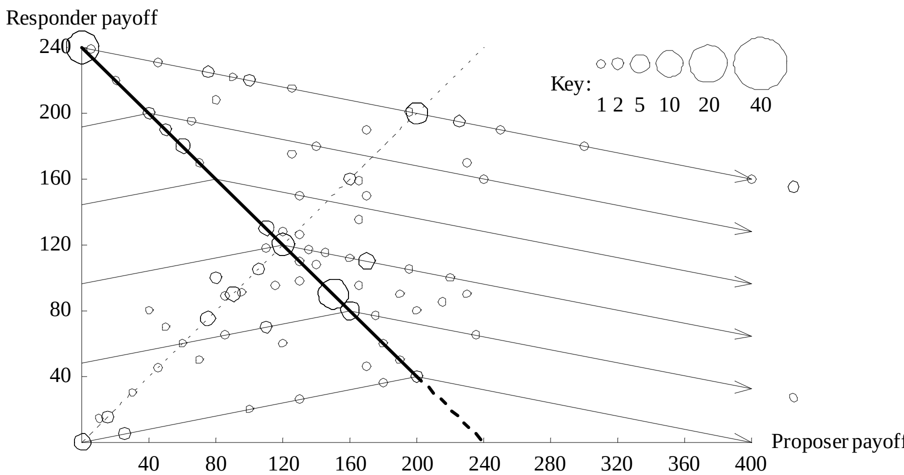

Figure 10. Outcomes in the Carrot-Stick Treatment (last 5 periods).

Figure 11 shows the outcomes in the Carrot treatment. Once again the substantial treatment difference is clear. Relative to the Carrot-Stick treatment we observe substantial rewards of low offers, and independent of the offer the rewards are much larger in the Carrot treatment. Here 35 percent of all outcomes are in area IV, which corresponds to 70 percent of the subjects making at least one choice which decreases both their private and relative payoffs. In addition, 23 percent of the subjects never make a change to the proposer's payoff over the 5 rounds.

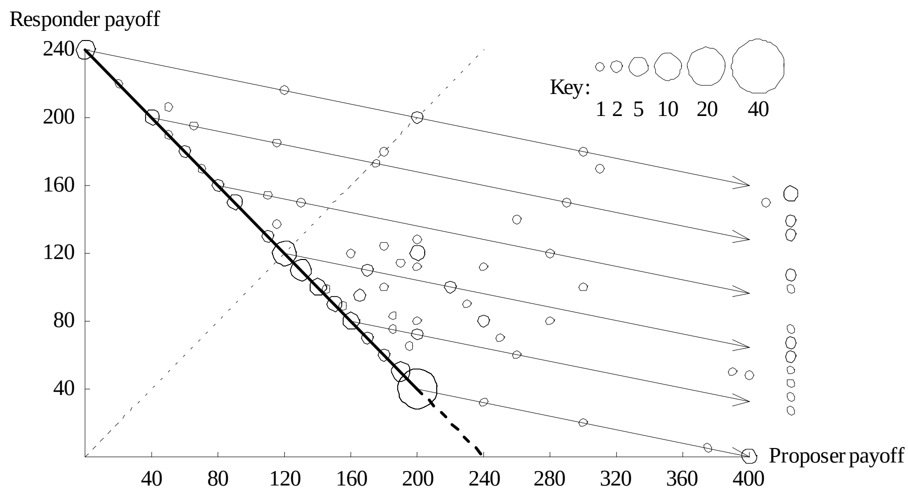

Fig 11: Outcomes in the Carrot Treatment (last 5 periods).

Finally, Figure 12 shows the outcomes in the Stick treatment. Here one third of the subjects chose at least one outcome in area II of the graph. In addition 43 percent of the responders chose no change to proposer's payoff.

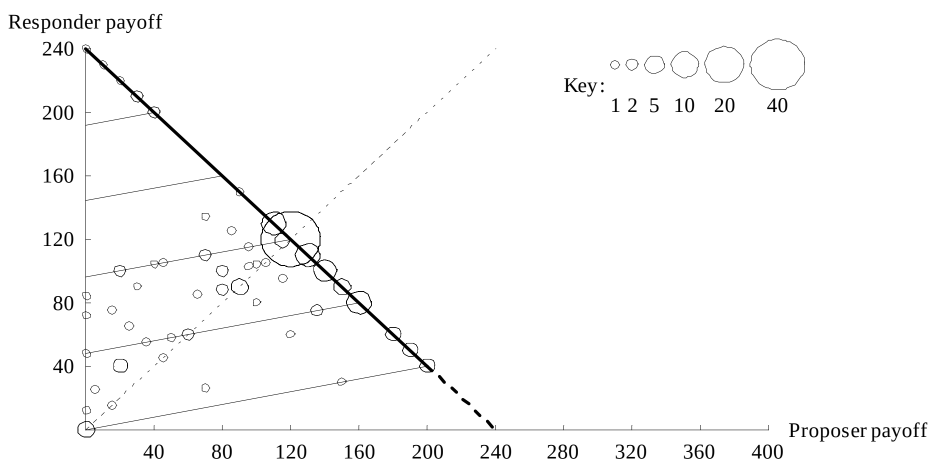

Figure 12. Outcomes in the Stick Treatment (last 5 periods).

To summarize the extent to which subjects are likely to choose outcomes in area's II and IV, we categorize them into one of three groups. The first group consists of subjects who never chose to change the proposer's offer. We refer to this group as payoff maximizers. The second group consists of those who either chose no change or a change that increased equality, i.e. these outcomes are in area I or III. We refer to this group as equalizers. The final group includes those who chose at least one outcome in area II or IV. The classifications are shown in Table 2 below.

Table 2: Number of Subjects in each Category

	Types		
	Payoff Maximizers	Equalizers I and III	Others II and IV
All ten rounds			
Carrot	6	1	23
Carrot-Stick	2	6	22
Stick	10	7	13
Last five rounds			
Carrot	7	2	21
Carrot-Stick	3	11	16
Stick	13	7	10

When only rewards are available, the vast majority of responders fall into the other category, however even in the Stick treatment we see that at least one third of the subjects fall into this category. The number of subjects who behave in a purely payoff-maximizing manner is largest when only punishments are available - given the vast differences in offers it is difficult to compare this to the number of payoff maximizers in the other treatments. It is interesting to note that about one third of the subjects choose equalizing outcomes.

5 Can Suggested Models Explain the Results?

Our experimental evidence is not consistent with the assumption that subjects have preferences that are defined solely over their private monetary payoffs. Rather it appears that individuals have preferences that also depend on the payoffs of others.

In the last decade a number of behavioral models have attempted to formalize these other-regarding preferences. The objective of this section is to determine whether this class of behavioral models can help shed light on the three puzzling findings of this experiment: the rewards of low offers; the larger rewards when there is no option of punishing; and the limited effect of the reward option on the demand for punishment.

Common to these behavioral alternatives is that individuals may compare the current distribution of payoffs to a reference point. If their current payoff standing is inferior to the reference point then they may wish to decrease the payoffs of others, and if their current standing is superior to the reference point then they may prefer to increase the payoffs of others. Subjects vary in their concern for the reference point, and individuals are faced with different trade-offs between maximizing their personal payoff and getting the preferred distribution of payoffs.

One class of models suggests that, independent of past interactions, subjects have a reference point of equal division of payoffs.[25] In particular, a subject with this motivation will equalize everyone's payoff to his private payoff if it is costless to do so. It is straightforward to derive predictions from this model since the reference point is independent of the availability of punishments or rewards. If we truncate the responses in the Carrot-Stick treatment by setting the rewards equal to zero, then these models predict that the conditional demand for punishments should be the same in the Stick and the Carrot-Stick treatments, i.e. $p_S = p_{CS}$. Similarly if we set all the punishments equal to zero then the demand for rewards should be the same in the Carrot and Carrot-Stick treatments, i.e. $r_C = r_{CS}$. While this prediction is correct for the punishments of medium sized offers, it is not consistent with the larger rewards in the Carrot treatment.

Furthermore when examining the distributions of final payoffs chosen by the responder it is clear that a large number of our subjects are not difference averse. If an equal division is desirable then subjects should never chose to change the proposer's payoff to a point where the final payoffs are in areas II and IV of Figure 3. The reason is that a responder in these areas could have chosen outcomes which simultaneously increase her payoffs and decrease the inequality in payoffs. Given an offer π_R a difference-averse response has the following characteristics: $r = 0$ when $\pi_R \leq 120$, and $r \leq \frac{5\cdot(\pi_R - \pi_P)}{6}$ when $\pi_R > 120$, similarly $p = 0$ when $\pi_R \geq 120$ and $p \leq \frac{4\cdot(\pi_R - \pi_P)}{5}$ when $\pi_R < 120$. Section 4.4 showed that between one-third and two-thirds of the responders choose outcomes that are inconsistent with these characteristics.

[25] Examples of difference-aversion models are presented in Loewenstein, Thompson, and Bazerman (1989), Bolton (1991), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000).

Finally, it is clear that difference aversion does not help explain the large rewards of small offers.

There is another behavioral alternative to difference aversion which may help explain the three surprising features in our data. The models of reciprocity suggest that subjects choose to increase or decrease the proposer's payoff because they want to reciprocate intentional actions. In particular an intentionally kind act may be rewarded and an unkind act punished. In the current reciprocity models the kindness of the proposer is defined relative to a reference point which depends on the largest and smallest Pareto efficient payoffs that the responder could have received.[26] While these models have the potential of predicting different reference points across different games, that is not the case here. Since the responder's set of Pareto efficient payoffs are the same across treatments, these models predict that the kindness associated with a particular offer is independent of the availability of rewards or punishments. Differences in demands for reciprocation therefore only depend on whether the responder's preferred action is perceived differently across treatments. Dufwenberg and Kirchsteiger (1999), extending Rabin (1993) to account for sequential games, predict that rewards and punishments are independent of one another and that there should be no differences across treatments.

Similarly the models have difficulty explaining rewards of low offers. The predictions in the intentionality models depend on the reference point at which an offer is defined as kind. However most of the currently used reference points would suggest that offers in our treatments need to exceed 120 to be considered kind.

A recent model by Charness and Rabin (2000) sheds some light on the increases of proposer's payoff after a small initial offer. They argue that subjects may have quasi-maximin preferences, and therefore may choose Pareto superior outcomes when they are available.[27] If the reference point is fixed across treatments, however, this model cannot explain the differences that we find in demands across treatments.

So the question remains - why are the demands for rewards affected by the availability of punishments, when punishments are not affected by the availability of rewards? There may be many different answers, yet it appears that only a model which allows for changes in the reference point across the three different treatments will be able to capture these differ-

[26] This type of reference point is consistent with that proposed by Rabin(1993) and Dufwenberg and Kirchsteiger (1998).

[27] A similar point is made in Andreoni and Miller (1998).

ences. The question is whether it is reasonable to assume that the reference point changes, in particular is it reasonable that the same offer be perceived differently across treatments.[28] It certainly appears as if different norms develop in the three games. Since both the average and modal offer is largest in the Carrot-Stick treatment, and offer may be perceived as less kind in the Carrot-Stick treatment compared to the two other treatments. If that is the case then we would expect larger rewards in the Carrot treatment and larger punishments in the Carrot-Stick treatment. Combined with the substitutability arguments this suggests that two opposing factors may be affecting the demand for punishments. On one hand the absence of rewards may cause a substitution towards the punishment option, and on the other hand the lack of rewards may imply that a given offer is perceived as being more generous suggesting less punishment. In contrast both of these effects suggest larger rewards when there is no punishment option.

6 Conclusion

We have examined the demands for rewards and punishments and their effects on cooperation. We considered a simple proposer-responder environment with randomly rematched partners. In this way, our experiment allowed us to concentrate on the pure demands for rewards and punishments. By considering three conditions—reward-only, punishment-only and reward and punishment—we were able to identify the effect each has separately as well as together.

We find, first, and perhaps surprisingly, that rewards alone are ineffective in generating generous offers. In the Carrot treatment the modal offer was the most selfish one possible, despite a willingness to reward by

[28] One way of thinking about a change in reference point may be the following. Picture yourself as a responder in the Carrot-Stick or the Stick treatment. Would you perceive an offer of say 40 the same way in the two treatments? In the Carrot-Stick treatment you may have contemplated a reward of a sufficiently high offer, hence it may be that a low offer is perceived as less kind in the Carrot-Stick than in the Stick treatment. This suggests that there may be larger average punishments in the Carrot-Stick treatment. Next let us consider the Carrot versus the Carrot-Stick treatment. Is there reason to perceive an offer of, say, 120 differently in the two treatments? One objective of making a large offer in the Carrot-Stick treatment is to avoid punishments, hence it may be that the same offer is perceived as kinder in the Carrot treatment where there is no such motive. Unfortunately the argument is circular, the reason is that it justifies differences in current rewards by assuming that rewards everywhere else are the same. In absence of a measure of the subjects' expectations, this justification for a change in reference points is therefore not satisfactory.

responders. Second, punishments improved cooperation but only eliminated extremely selfish offers, pushing proposers in the Stick treatment to modest degrees of cooperation. Combining rewards and punishments had a very strong effect. In this Carrot-Stick treatment the modal offer was the most generous one possible, often leading to rewards by responders. Even though generous offers were not punished, such generosity was only reached when the threat of punishments existed. This indicates that rewards and punishments act to complement each other and, even though only one can be used at a time, the combination of the tools leads to the greatest degree of cooperation.

In addition to this, we also found some surprising treatment effects in the responders' demands for rewards and punishments. While demands for punishments are unaffected by the availability of a reward option, we found that rewards are larger when there is no option of punishing. From the perspective of current theoretical models of fairness, the combination of these two findings is quite puzzling and does not fit the models. It appears that an explanation may require that the definition of kindness changes systematically by treatment. Given the distribution of offers it may be that a particular offer is perceived as less kind in the Carrot-Stick treatment, than in the two other treatments. This result suggests many productive areas for future research, both theoretical and experimental.

Finally, what do our results suggest about how fairness may shape economic institutions? While more work clearly needs to be done, this indicates that cooperation is most successfully enforced in an environment in which both punishments and rewards are allowed. The process suggested by our data is that the stick can help by getting people to move away from perfect selfishness and to test the waters of cooperation. The carrot can then take over by encouraging further cooperation, rendering the stick a rarely used but important and necessary tool.

References

Andreoni, James and John H. Miller, "Giving According to GARP: An Experimental Test of the Rationality of Altruism," working paper, University of Wisconsin, 1998.

Andreoni, James, Paul Brown, and Lise Vesterlund, "What Makes an Allocation Fair? Some Experimental Evidence," Unpublished Manuscript, 1999.

Andreoni, James, Marco Castillo, and Ragan Petrie, "What do Bargainers' Preferences Look Like? Exploring a Convex Ultimatum Game," Unpublished Manuscript, 2000.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, Reciprocity, and Social Norms," *Games and Economic Behavior*, 1995, 10, 122-142.

Binmore, Ken, Avner Shaked and John Sutton, "Testing Non-cooperative Bargaining Theory: A Preliminary Study," *American Economic Review,* 75, 1985, 1178-80.

Bolton, Gary E. and Rami Zwick, "Anonymity versus Punishment in Ultimatum Bargaining," *Games and Economic Behavior*, 10, 1995, 95–121.

Bolton, Gary E., "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, 81,1991, 1096-1136.

Bolton, Gary E., and Axel Ockenfels, "ERC, A Theory of Equity, Reciprocity and Competition," *American Economic Review*, 90(1), 2000, 166-193.

Charness, Gary, and Matthew Rabin, "Some Simple Tests of Social Preferences," Unpublished Manuscript, 2000.

Charness, Gary, "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation," Unpublished Manuscript, 1996.

Charness, Gary and Ernan Haruvy, "Altruism, Equity, and Reciprocity in a Gift-Exchange Experiment: An Encompassing Approach," forthcoming *Games and Economic Behavior.*

Charness, Gary, Ernan Haruvy, and Doron Sonsino, "Social Distance and Reciprocity: An Internet Experiment," Unpublished Manuscript, August 2000.

Cox, James C. "Trust and Reciprocity: Implications of Game Triads and Social Contexts," Unpublished Manuscript, 2000.

Dickinson, David, "The Carrot vs. the Stick in Work Team Motivation," Unpublished Manuscript, 1999

Dufwenberg, Martin and Georg Kirchsteiger, "A Theory of Sequential Reciprocity," Unpublished Manuscript, 1998

Dufwenberg, Martin and Uri Gneezy, "Measuring Beliefs in an Experimental

Lost Wallet Game," *Games and Economic Behavior*, 30, 2000, 163-182.

Eckel, Catherine C. and Philip J. Grossman, "The Relative Price of Fairness: Gender Differences in a Punishment Game," *Journal of Economic Behavior and Organization,* 30(2), 1996, 143-158.

Eckel, Catherine C., and Philip J. Grossman, "Altruism in Anonymous Dictator Games," *Games and Economic Behavior*, 16, 1996, 181–191.

Falk, Armin and Urs Fischbacher, "A Theory of Reciprocity," Unpublished Manuscript,1999.

Fehr, Ernst and Simon Gächter, "Cooperation and Punishment in Public Goods Experiments," forthcoming in *American Economic Review.*

Fehr, Ernst, and Klaus M. Schmidt, "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114, 1999, 817-868.

Forsythe, Robert, Joel Horowitz, N.S. Savin, and Martin Sefton, "Fairness in Simple Bargaining Games," *Games and Economic Behavior*, 6, 1994, 347–69.

Güth, W., R. Schmittberger, and B. Schwarze, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization,* 3, 1982, 367-388.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler, "Fairness and the Assumptions of Economics," *Journal of Business,* 59, 1986, 285–300.

Loewenstein, George, Leigh Thompson, and Max Bazerman, "Social Utility and Decision Making in Interpersonal Contexts," *Journal of Personality and Social Psychology,* 57, 1989, 426-441.

McKelvey, Richard D. and Thomas R. Palfrey, "An Experimental Study of the Centipede Game," *Econometrica*, 60, 1992, 803–836.

Offerman, Theo, "Hurting Hurts more than Helping Helps," Unpublished Manuscript, 2000.

Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1993, 1281–1302.

Rabin, Matthew, "Bargaining Structure, Fairness, and Efficiency," Unpublished Manuscript, 1997

Sefton, Martin, Robert Shupp, and James Walker, "The Effect of Rewards and Sanctions in the Provision of Public Goods," Unpublished Manuscript, 2000.