

January, 2008

Social Preferences and Moral Biases

Rachel Croson

Director, The Negotiations Center
University of Texas at Dallas
800 W. Campbell Drive
Richardson, TX 75080-3021
crosonr@utdallas.edu

James Konow*

Department of Economics
Loyola Marymount University
One LMU Drive, Suite 4200
Los Angeles, CA 90045-2659
jkonow@lmu.edu

Abstract

A consensus seems to be emerging in economics that at least three motives are at work in many strategic decisions: distributive preferences, reciprocal preferences and self-interest. An important obstacle to this research, however, has been *moral biases*, i.e., the distortions created by self-interest that can obscure our measures of social preferences. Among other things, this has led to disagreement about the relative importance of self-interest, distributive and reciprocal preferences. This paper describes a simple experiment that decomposes behavior into these three forces. We compare the decisions of implicated “stakeholders” with those of impartial “spectators,” who have no stake. Several surprising and interesting results emerge. For example, stakeholders are less inclined to respond to the generosity of others than are spectators acting on their behalf. This experiment also helps clarify a result in previous research (e.g., Offerman, 2002) that stakeholders tend to punish unkindness more than they reward kindness. We find that this asymmetry in reciprocity has two sources: there is an asymmetry in the underlying preference that even impartial spectators display, but, in addition, stakeholders exhibit a moral bias, i.e., they punish more and reward less than spectators. In sum, we find that all three motives have important and significant effects on final allocations.

Keywords: Reciprocity, fairness, justice, moral bias

JEL classification: D63, C91

*Corresponding author. We thank the editor and two referees of this journal, Alexander Cappelen, Gary Charness, Simon Gächter, David George, Bertil Tungodden, participants of seminars at the Norwegian School of Economics and Business, Notre Dame University, the University of Oslo, and at meetings of the Allied Social Science Association, Economic Science Association, Asia Pacific Economic Science Association, and Public Choice Society for helpful comments and suggestions. Croson acknowledges support from NSF SBR-9876079. We assume all responsibility for any remaining errors.

Nature, which formed men for that mutual kindness, so necessary for their happiness, renders every man the peculiar object of kindness, to the persons to whom he himself has been kind. Though their gratitude should not always correspond to his beneficence, yet the sense of his merit, the sympathetic gratitude of the impartial spectator, will always correspond to it.

Adam Smith, *The Theory of Moral Sentiments*

1. Introduction

The assumption of self-interest has served as a powerful axiom of economics. It has helped to explain and predict large swaths of observed facts and to develop rigorous and elegant theoretical models. With mounting evidence of behavior at variance with material self-interest, however, economists are increasingly enriching traditional models with additional motives. The growing consensus is that integrating *social preferences* into the analysis can explain important economic phenomena, including involuntary unemployment (e.g., Akerlof and Yellen, 1990), pricing policies (e.g., Kahneman, et al., 1986b, Kachelmeier, et al., 1991) and bargaining behavior (e.g., Güth, Schmittberger and Schwarze, 1982).

Nevertheless, the confluence of self-interest and social preferences has proven an important hurdle to inferring the underlying preferences that theory needs to incorporate. Attempts to identify social preferences have been hindered by what we call “moral biases,” i.e., the distortionary effects of self-interest on expressed social preferences. For example, consider the “dictator game,” i.e., an experiment in which one subject (the dictator) may transfer money to an anonymous recipient. A purely self-interested dictator would send nothing. On average, though, dictators send approximately 30% of their endowment to recipients. This 30% transfer is typically taken as a measure of social preference. But suppose that an impartial spectator motivated purely by social preferences would implement an equal split between the parties. We take this as the measure of true social preferences. The observed 30% transfer is neither self-interest, which implies no transfer, nor social preferences, which prescribe a 50% transfer. The difference between the 30% transfer and the 50% transfer is an

example of what we call a moral bias; it represents a distortion in our measures of social preferences due to the impact of self-interest on their expression.¹

Although the existence of moral biases has been noted (e.g., Forsythe et al., 1994), almost no studies of social preferences have isolated social preferences from self-interest and its distorting effects. One exception is Charness and Rabin (2002), who report two sets of comparisons in binary choice dictator games focusing on preferences for surplus maximization. The current study, on the other hand, has fixed stakes and sets out to explore distributive and reciprocal preferences over a wide range of allocation conditions. We experimentally isolate the separate effects of self-interest, distributive and reciprocal preferences using a simple two-stage dictator game. We contrast the behavior of *stakeholders*, or implicated parties, with that of *spectators*, or third parties, the latter serving to establish a benchmark of pure social preferences.

Our design permits us to identify and separate the different motives. The experiment produces evidence of moral bias that affects both the average willingness to act on social preferences as well as responsiveness to inequity, kindness and unkindness. In addition, both spectators and stakeholders exhibit an asymmetry in reciprocity, favoring punishment over reward. There is also a moral bias in this asymmetry: spectators are more inclined than stakeholders to reward and less inclined to punish.

The paper is organized as follows. Section 2 presents background and hypotheses, section 3 describes the experimental design and procedures, section 4 presents and analyzes

¹ We should distinguish moral bias from another phenomenon that has been noted in this literature called self-serving bias. *Moral bias* can be traced to at least two sources: *self-serving bias* and *self-centered bias*. As commonly used in the literature, self-serving bias refers to an alteration of one's beliefs (self-deception) due to self-interest, e.g., believing it is fair to be unfair in order to relieve the disutility of engaging in unfair behavior. In contrast, the self-centered bias involves deliberately self-interested action, e.g., unfair behavior that is acknowledged by the actor to be unfair and yet is still chosen. We will not distinguish these two forces in this paper but instead will use the more general term, moral bias.

the results of the experiment, and section 5 contains the concluding remarks.

2. Background and Hypotheses

2.1 Background: Social Preferences and Experimental Design

This subsection selectively reviews previous experimental studies of social preferences in order to clarify and motivate our experimental design.

The Güth et al. (1982) experimental test of the ultimatum game (and many follow-up studies) demonstrated that behavior deviated from that predicted by self-interested subgame perfect equilibria. These results framed much of the subsequent literature on social preferences. Among possible explanations for these and similar “anomalies,” authors offer fairness (e.g., Bolton 1991, Bolton and Ockenfels 2000, Fehr and Schmidt 1999), altruism (e.g. Becker 1974, Levine 1998), warm glow (e.g. Andreoni, 1993), spite (e.g. Cason, Saijo and Yamato, 2002), intentions (e.g. Rabin, 1993), and trust and reciprocity (e.g. Berg, Dickhaut and McCabe, 1995). The central debate has identified two types of social preferences: *distributive*, i.e., preferences over outcomes or endstates, and *reciprocal*, i.e., preferences over intentions or player types.²

The experimental ultimatum design of Blount (1995) clarifies this distinction and motivates our design. Suppose a proposer offers a responder \$3 out of a sum of \$10. Now compare this with the same division, but suppose it had been randomly determined. If the responder’s preferences were solely distributive, she would either accept or reject this offer, regardless of its source. If, on the other hand, the responder were (also) motivated by reciprocity, she might reject the proposer’s offer as unkind but accept the randomly

² We should note that terminology is evolving with the accumulation of knowledge in this area, and sometimes the debates over what to call these preferences have run hotter than the debates over the actual form of the preferences themselves. We will be more specific in our meaning later, but whatever usage one adopts, it is bound to conflict with that used in some important part of this literature.

determined \$3. We adopt this feature of Blount's, comparing responses to divisions made by one's counterpart with random divisions. In addition, our design eliminates explicit strategic elements and allows for both negative and positive reciprocity.³

Another important design in this literature has been the trust (or investment) game, introduced by Berg, Dickhaut and McCabe (1995). Proposers transfer money to responders, which is multiplied. The trust game produces efficiency gains, unlike the ultimatum or dictator games. Responders can then return some, all or none of the amount received to proposers. Contrary to the subgame perfect equilibrium, proposers typically send some positive amount to responders, who then often return positive amounts to proposers (e.g., see Croson and Buchan 1999). The latter is sometimes taken as a measure of reciprocity and the former of trust. However, both proposer and responder transfers might be motivated by distributive preferences, as well. Cox (2004) introduced a "triadic" design that supplements the standard trust game experiment with two treatments that are variations on the dictator game.⁴ The results show significant evidence of distributive preferences, as well as evidence of trust and reciprocity.

Previous studies have allowed expression of both positive and negative reciprocity, including Abbink, Irlenbusch and Renner (2000), Charness (2004) and Offerman (2002), and have found that negative reciprocity appears to be a more powerful motivation than positive

³ Brandts and Solà (2001) and Falk, Fehr and Fischbacher (2003) employ a variation called the "mini-ultimatum game" that produces strong evidence of negative reciprocity: the rejection rate by responders to a given proposer offer depends on the alternate offer a proposer could have made.

⁴The second treatment is the same as the standard trust game, but the recipients cannot send anything back to proposers. In the third treatment, each person in one group is endowed with the same amounts kept by proposers in the standard trust game, and each person in the second group is endowed with the tripled amounts received by responders in that game, after which each person in the second group may transfer any amount to her counterpart in the first group. Trust is interpreted as any change in proposer transfers between the first and second treatments, whereas reciprocity can be seen as any difference in responder transfers between the first and third treatments. McCabe, Rigdon and Smith (2003) introduce another clever variation on the trust game that further supports the importance of trust and reciprocity. One reader, however, comments that these designs still do not explicitly account for any role for risk aversion.

reciprocity. These papers focus on the reciprocal preferences of self-interested stakeholders. Our design, in contrast, examines the relative strength of positive and negative reciprocity in stakeholders as well as in third party spectators.

Recent models have incorporated the three forces of distributive, reciprocal and self-interested motives, e.g., Charness and Rabin (2002), Cox, Friedman and Gjerstad (2004) and Falk and Fischbacher (2000). Our study decomposes the motivational forces behind allocation decisions into these three parts.⁵ The challenge, which we believe our design overcomes, is to decompose these forces without the distortions that potentially plague such measures when multiple motives are activated. Although previous studies have examined these motives, we believe this is the first to decompose and analyze them.

Previous work has investigated third party preferences. For example, Kahneman, Knetsch and Thaler (1986a) conducted an experiment, akin to a binary choice version of the dictator game, in which subjects are willing to forgo \$1 to punish a subject who had been unfair to someone else in a previous decision. In the public goods experiment of Carpenter and Matthews (2004), many subjects are willing to sanction players in other groups, even at a cost. Fehr and Fischbacher (2004) report the results of a series of experiments demonstrating third party punishment when distribution and cooperation norms are violated. These studies provide compelling evidence of the willingness of third parties to punish. They do not, however, answer the particular set of questions we wish to address here. First, their third parties are stakeholders rather than spectators, i.e., they must incur a cost to punish.⁶ This underscores the strength of the reciprocal preference, but it also opens the possibility of a

⁵ The data produced by the current study permit quantification of these motives in dollar terms but not estimation of a random utility function.

⁶ It might be that having a stake triggers an emotional response missing from our spectators. But our aim in this study is precisely to separate all such stakeholder considerations from spectators.

moral bias in its measurement.⁷ Second, they focus on negative reciprocity, whereas we examine, for both second and third parties, social preferences of all types: distributive, positively reciprocal and negatively reciprocal. Finally, first-stage dictators in our study are not informed of the second stage until it is reached in order to minimize or eliminate strategic considerations. This design feature is lacking in most previous studies.

Some experiments have revealed an efficiency motive, as in Andreoni and Miller (2002) and Charness and Rabin (2002), or a sense of desert from earned amounts, as in Cherry, Frykblom and Shogren (2002), Hoffman et al. (1994), Konow (2000), and Rutström and Williams (2000). In order to simplify our task, we eliminated these considerations. Thus, the stakes in both stages of our experiment are fixed and are endowed rather than earned.

We believe a significant obstacle to identifying social preferences is the presence of moral biases. Self-interest appears not only in unmistakable forms but can also mask itself as social preferences when strategic interaction is possible. Consider the ultimatum game. Even a purely self-interested proposer might make a generous offer, as small offers are likely to be rejected, suggesting that ultimatum game offers *overstate* proposers' true distributive preferences (consistent with the results of Forsythe et al., 1994). Strategic behavior can also confound inferences about reciprocal preferences in the trust game. Senders who anticipate positive reciprocity have a self-interested incentive to transfer money. Thus, responders do not know whether proposer generosity is due to the sender's social preferences, which she would like to reward, or to his self-interest, which she might rather punish. Hence, responder transfers might *understate* their reciprocal preferences.

⁷ These previous studies also help dismiss any concern that third party decisions fail to have salience, in the terminology of experimental economics. Indeed, unless one rejects social preferences altogether, third party decisions should be quite salient. If they were not, one would probably expect very simple or random decisions. As we later report, however, third party decisions are both sophisticated and subject to less unexplained variance than those of stakeholders.

In light of this, we eliminate strategic interaction in our experiment. We use a dictator game involving the allocation of \$10 between two subjects in the first stage, followed by an *unannounced* second stage in which a subject allocates \$20 between the two subjects from the first stage. One important variable concerns the identity of the second-stage allocator. In one set of treatments it is a stakeholder, viz., the recipient from the first stage, whereas in the other it is a spectator, i.e., a third party, who is paid a fixed fee unrelated to the allocation she chooses. We compare, therefore, the decisions of an impartial spectator, whose actions reflect unadulterated social preferences, with those of an implicated stakeholder, whose decisions might also reflect self-interest.

2.2 Hypotheses

In light of previous work and our design, we propose the following hypotheses.

H1. There is a moral bias in overall social preferences.

That is, when both distributive and reciprocal preferences are relevant, stakeholders are less willing than spectators to act on them. Here we seek to replicate this finding from other studies and additionally to analyze the effects of self-interest on both the average level of and degree of responsiveness in social behavior.

H2: There is a moral bias in distributive preferences.

Stakeholders are less willing than spectators to act on distributive preferences, both on average and in their responsiveness to inequities of different magnitudes.

H3: Stakeholders have asymmetric reciprocal preferences: they are more likely to punish (negative reciprocity) than to reward (positive reciprocity).

Offerman (2002) previously identified this asymmetry and speculated that it is due to a self-

serving bias (a type of moral bias) on the part of stakeholders that motivates stakeholders to rationalize punishing unkindness but not rewarding kindness. By introducing spectators, we are able to examine whether this is the case or whether the asymmetry is characteristic of the underlying social preference, as proposed in Hypothesis 4.

H4: Spectators have asymmetric reciprocal preferences: they are more likely to punish (negative reciprocity) than to reward (positive reciprocity).

Even if both spectators and stakeholders are found to exhibit this asymmetry (or neither group displays it), that does not rule out a moral bias in reciprocity. There might be differences in the degree to which stakeholders reward or punish relative to spectators that is indicative of a moral bias. For example, even if we fail to reject both H3 and H4, we might still find that spectators are more likely to reciprocate positively (and less likely to reciprocate negatively) than are stakeholders. We propose this in our final hypothesis.

H5: Spectators are more likely to reward and less likely to punish than stakeholders.

3. Description of the Experiment

3.1 Experimental Design

The experiment is a two stage dictator game. In the first stage, \$10 is distributed between each of two paired subjects denoted X and Y. One treatment variable has to do with the method of this distribution. In the “X Decision” conditions, each subject in the X group receives \$10, which he can divide in even dollar amounts, as dictator, between himself and an anonymous counterpart in group Y. In the “Random Division” condition, the \$10 sum is divided in one of the same six possible ways between each subject in groups X and Y, but the exact division is randomly chosen for each pair.

In the second stage, which is not previously announced to subjects, a sum of \$20 is divided dictator-fashion between each of the X, Y pairs. The second treatment variable concerns the identity of this second-stage dictator (here called the allocator). In the “Stakeholder Y Allocator” conditions, each Y subject chooses how much to allocate to herself and her X counterpart from the first stage in any one-dollar increment. The “Spectator Z Allocator” conditions are similar, except there is a third group, Z. Each subject in that group is assigned an X, Y pair and chooses the allocation of \$20 between these two subjects. This Z allocator receives a separate fixed \$20 fee for this decision that does not depend in any way on the allocation.⁸

We use the strategy method for all treatments: Stakeholder Y (or Spectator Z) chooses how much of the \$20 to give to X and Y for each of the six possible first-stage divisions. Thus, second-stage allocators make their choices without knowing what divisions occurred in the first stage. In both stages, we deliberately omit contextual elements that might activate preferences for efficiency, endowment, status, etc., in order to simplify the distributive task and, in that way, be better able gauge its impact relative to self-interest and reciprocity.

Figure 1. Experimental Design

		First-stage Method	
		Random Division	X Decision
Second Stage Allocator	Stakeholder Y Allocator	R Y (X, Y)	D Y (X , Y)
	Spectator Z Allocator	R Z (X, Y, Z)	D Z (X , Y, Z)

Note: Subjects in the role of decision maker in a given treatment are indicated in bold.

These treatment variables are crossed to yield a 2 First-stage Method (X Decision,

⁸ Engelmann and Strobel (2004) similarly seek to extract self-interest but do so by holding constant the dictator’s payoff across a discrete set of allocations between the dictator and two other subjects.

Random Division) \times 2 Second-stage Allocator (Stakeholder Y Allocator, Spectator Z Allocator) factorial design. Thirty pairs or triples, respectively, participated in each cell in an across-subjects design. Figure 1 summarizes the experimental design and the abbreviations we will use for these treatments. Letters in bold represent the decision-making roles in the treatments.

We use the strategy method for second-stage allocations for several reasons.⁹ First, it allows us to observe distributive and reciprocal preferences at the individual level.¹⁰ Second, it provides a richer set of observations than would be possible using only actual first-stage allocations, since certain divisions are seldom chosen by first-stage dictators. This was an especially acute concern in the case of this experiment, since we sought data on positive reciprocity (i.e., rewards) in response to X Decisions that give Y more than one-half, and such decisions are very rare. We addressed concerns about the strategy method in part by simplifying the cognitive task, e.g., by using straightforward and clear wording and procedures. This was also the reason for constraining possible first-stage divisions to even dollar amounts: second-stage allocators needed to make only six allocations as opposed, say, to eleven had first-stage divisions been in any dollar amount.

Of the four treatments, treatment DY is closest in spirit to the standard trust game.

Two important differences here are fixed stakes (transfers are not multiplied)¹¹ and the

⁹ It is possible that the strategy method elicits different behavior in comparison to responses to actual decisions. This issue has not yet been resolved in the literature: Brosig and Weimann (2003) do find significant differences between these methods, but Brandts and Charness (2000) and Cason and Mui (1998) do not. One reader pointed to evidence elsewhere that the emotions of stakeholders are activated with single decisions in such circumstances but that such activation is not clear in the case subjects choose a schedule of actions. That might be, but we still find significant reciprocal behavior here and, therefore, would expect to find an even stronger effect, if anything, without the strategy method.

¹⁰ This provides evidence on a variety of questions, including whether subjects differ in their cut-off point between positive and negative reciprocity. Interestingly, we find differences in cut-off points are usually small, so we will not focus on them in later data analysis.

¹¹ In this respect, this treatment is similar to the Specific Reciprocity treatment of Ben-Ner, et al. (2004). Some differences between this study and that one include the fact that their experiment lasted about 2 hours, second-stage allocators made only one choice about what to send after finding out what their first-stage counterpart had sent, and

absence of strategic play (in the first stage, X subjects are not informed of the possibility of a second stage, and this fact is common knowledge).¹² Second-stage allocators can, therefore, take first-stage divisions as genuine measures of the willingness of X subjects to sacrifice material self-interest for social preferences, undistorted by X's strategic self-interest.

Another difference is the wide berth given to second-stage allocators to express both positive and negative reciprocity. Assume, for the sake of illustrating this point, that fairness call for equal splits of first and second-stage amounts. Then the second-stage allocator can punish X selfishness (when X retains more than half of the \$10 for himself) by withholding part or all of X's \$10 entitlement to one-half of the \$20 second-stage earnings. Similarly, the second-stage allocator can reward X generosity (when X retains less than one half of the \$10 for himself) by bestowing more than one-half of the second-stage earnings, between \$10 and \$20, on X. Thus, second-stage allocations in the DY treatment reflect the confluence of three potential stakeholder Y motives: *self-interest*, *distributive preferences* and *reciprocity*, as in the trust game, except that the current design provides evidence of X's social preferences without strategic or efficiency motives.

This is depicted in Figure 2 below, which summarizes the types of preferences that operate in each treatment. By examining the differences between these treatments we can isolate the various types of preferences in operation.

The RZ treatment is the most basic one in terms of motivation. In this treatment, X subjects make no decisions in the first stage, since the initial division of the \$10 is random.

second-stage decisions were over the same \$10 amounts as first-stage decisions. Our study also involves the three other treatments detailed above.

¹² The information provided to subjects is, therefore, always truthful, even if it is, in some respects, incomplete. This is similar to the "restart" design of Andreoni (1988) and Croson (1996), the use of dictator decisions to screen players for an unannounced second game in Charness (2000), and the "double dictator" treatment in Konow (2000).

Second-stage allocators, therefore, have no basis for reciprocity, since X subjects are unable to express their social preferences. The second-stage allocators are spectators (Z subjects), who have no stakes and whose decisions, therefore, are not biased by self-interest. The RZ treatment serves, then, to calibrate the pure effect of distributive preferences. It also provides a baseline for quantifying the effects of other motives.¹³

Figure 2. Types of Preferences and Treatments

		First-stage Method	
		Random Division	X Decision
Second Stage Allocator	Stakeholder Y Allocator	R Y Self Interest, Distributive	D Y Self Interest, Distributive, Reciprocal
	Spectator Z Allocator	R Z Distributive	D Z Distributive, Reciprocal

Two comparisons reveal the aforementioned moral biases. In the RY and RZ treatments, initial allocations are random, and second-stage allocators can act on their distributive, but not reciprocal, preferences. Thus, one can measure the effect of self-interest in the presence of distributive preferences by comparing the differences in second-stage allocations between treatment RY, where self-interest and distributive preferences are potentially in play, and treatment RZ, which reflects only distributive preferences. One can similarly isolate the effect of self-interest when both distributive and reciprocal preferences are potentially implicated by comparing treatments DY and DZ treatments. In both

¹³ One might wonder why we test for distributive preferences in the RZ treatment instead of simply assuming, as seems reasonable in a contextually simple experiment, that impartial distributive preferences reduce to equal splits of the total amount from both stages. Indeed, equal split preferences are evident in the results of many contextually lean experiments, e.g., ones with a single stage of decision-making, anonymity, unearned endowments, fixed stakes, and no information about need, merit, gender, etc. On the other hand, other experiments reveal patterned deviations from equality, e.g., Babcock, et al. (1995), Gächter and Riedl (2001), and Konow (2000). Although the current experiment is contextually simple, some design characteristic, such as two stage decision-making, could prime preferences for unequal splits of the total. For example, even subjects who prefer equality might compartmentalize decisions in the two stages, i.e., second-stage decisions might tend toward equality but not entirely adjust for inequalities created in the first stage. Thus, we test for each type of social preference in this study, rather than imposing any specific assumption.

treatments, X subjects decide on the first-stage division (potentially activating both the distributive and reciprocal preferences of second-stage allocators), but any difference between the two indicates the effect of stakeholder Y self-interest (versus spectator Z impartiality), similar to the comparison above of treatments RY and RZ.

The effect of reciprocity can also be identified by two comparisons of treatments. First, second-stage allocators in the RZ and DZ treatments are spectator Z subjects with no personal stakes. In the RZ treatment, these spectators can express only distributive preferences given the random determination of first-stage allocations, whereas in the DZ treatment first-stage divisions differ because of X's decision, potentially activating both the distributive and reciprocal preferences of second-stage allocators. Differences in these treatments, therefore, should reveal the pure effect of reciprocal preferences in spectators. Second, the RY and DY treatments parallel the RZ and DZ treatments, respectively, in terms of how the initial \$10 is divided. In the RY and DY treatments, however, second-stage allocators are stakeholders Y rather than spectators Z, introducing a role for self-interest. Thus, a comparison of these treatments reveals the effect of reciprocity (i.e., of stakeholders), when both self-interest and distributive preferences are relevant.

3.2 Experimental Procedures

After registering, receiving their show-up fee and being seated (randomly, in the case of X/Y sessions), subjects received a form with the first-stage instructions and allocation information. The experimenter then read the instructions aloud. The experiment was run on paper and conducted single-blind: each subject is identified only by a subject ID, which only he and the experimenter knew. Subjects were told that they would never know the identity of their counterparts. All subjects were provided with the same information, but no one was

informed, at this point, of the second stage. In the X Decision conditions, X subjects chose one of the six X/Y divisions by circling it. In the Random Division condition, one of the lines was already circled on the X forms only.¹⁴ These forms were then collected.

Now subjects were told for the first time of a second stage of the experiment and were informed that this was the final decision. They received forms with the second-stage instructions and space, if applicable, for decisions. In RY and RZ sessions, subjects were informed that the \$10 stakes in the first stage had been randomly divided between X and Y subjects. In the DY and DZ sessions, they were told that X had made this decision. Then, the second-stage allocator (stakeholder Y in the RY and DY sessions and spectator Z in the RZ and DZ sessions) chose how to allocate the \$20 between X and Y for each of the six possible first-stage divisions. These forms were then collected, and, while payments were calculated, subjects completed an exit questionnaire that asked demographic information and reasons for the decisions. Subjects turned in these forms, signed for payments and were free to leave.

Three hundred undergraduates participated in this experiment with thirty subjects per session.¹⁵ Subjects were randomly assigned to X or Y roles in X/Y sessions; Z sessions were conducted separately with only Z subjects. All decision-making sessions were conducted over two consecutive days, and all second-stage allocations were scheduled for the first day to avoid contamination effects. Two sessions were conducted after these two days, and these both involved X and Y subjects in the RZ treatment who made no decisions. Total average

¹⁴ The actual random transfer of payoffs used the full range of possible (x, y) transfers, $\{(10, 0), (8, 2), \dots, (0, 10)\}$. It was from a symmetric distribution that did not, therefore, favor X or Y, and that gave a higher probability to more equal transfers, as is common with actual X decisions. Specifically, the frequency of an allocation with \$x in these treatments was $f = .5 \cdot (19 - 3 \cdot |x - 5|)$, which produced a simple, piecewise linear distribution with these attributes.

¹⁵ There was one exception: one group of thirty X and Y subjects in the RY treatment was conducted over two sessions due to an unexpectedly large number of no shows in the first session. Most participants were recruited via e-mail and posted notices around campus to register at a website. A small number was recruited from a campus subject pool, which also satisfied class credit – these subjects were all assigned to non-decision-making roles (e.g., X in Random Division treatments or Y in Group Z allocator treatments).

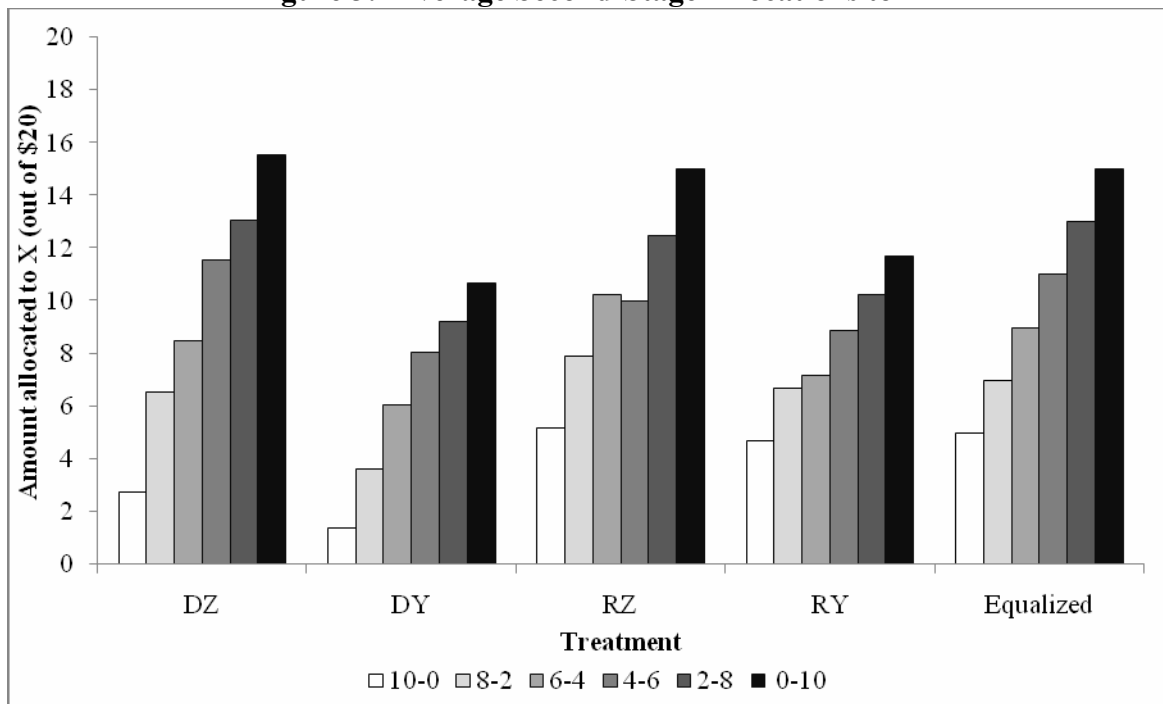
earnings per subject equaled \$21, which included a \$5 show up fee. Sessions lasted about 40 minutes, so that average hourly earnings were about \$32. At the end of the experiment, 94% of subjects indicated they would be willing to participate in an economics experiment again.

4. Results and Analysis

4.1 Summary of Results

Our primary interest is in the allocations made in the second stage of decision-making. We note the average first-stage division to Y in the X Decision sessions was \$3.07 (st. dev. 1.92), or around 30% of the available amount. These divisions are comparable to those in other dictator studies, which suggests first-stage decisions were untainted by expectations of a second stage.

Figure 3: Average Second-Stage Allocations to X



Second-stage allocator decisions were elicited using the strategy method, i.e., for each possible first-stage division. Figure 3 displays these second-stage allocations, where each set of bars represents one of the four treatments. The height of each bar shows the average

amount allocated to X corresponding to each of the six possible first-stage divisions. It is apparent from the increasing bar heights that the mean amount allocated to X in the second stage is positive monotonic in the mean amount received by Y in the first stage (with a single minor exception at 4-6 in the RZ treatment). For comparison, the final set of bars denoted “Equalized” describes the allocations to X needed in order to equalize earnings.

We compare second-stage allocations across treatments in two ways. First, we examine differences in the *level* of allocator generosity in the second stage. This involves calculating the *average* of the amounts allocated to subject X for each of the potential first-stage divisions, that is, the average within each treatment of the bars in Figure 2. Second, the fact that the second stage was implemented using the strategy method enables us to analyze differences in the *responsiveness* of second-stage allocators to first-stage divisions. In particular, we know the amounts the second-stage allocator (stakeholder Y or spectator Z) allocated to X for each of the six possible divisions of the \$10 in the first stage. From these we calculate five differences between the six first-stage divisions, and then derive the average difference from these for each second-stage allocator. Across all allocators, this can be interpreted as the average slope of the bars in each treatment in Figure 3.¹⁶

Table 1 presents summary statistics on averages and slopes of the second-stage allocations for the four treatments. Note that, when stakeholder Y is the second-stage allocator, the averages are significantly less than the \$10 that would be needed to equalize totals, both in the Random Division treatment ($p=.0005$) as well as in the X Decision treatment ($p<.0001$), according to two-tailed t-tests. The average amounts allocated to X by spectator Z allocators, on the other hand, are only marginally different from \$10, both in the

¹⁶The slope of these bars is the slope of the average amount sent over each individual. To calculate the slope for statistical purposes, we calculate the slope of each individual, and then average these over the individuals.

Random Division treatment ($p=.0965$) and in the X Decision treatment ($p=.0667$), again using two-tailed t-tests. Another difference between stakeholder Y and spectator Z allocators is that the standard deviations are significantly greater in the stakeholder Y treatments than in the spectator Z treatments (using a two-tailed F-test of variances, pooled Ys > pooled Zs $F=24.38$, $p=.0001$; $DY > DZ$ $F=9.92$, $p=.0026$; $RY > RZ$ $F=10.29$, $p=.0022$). These results on means and standard deviations are consistent with evidence elsewhere (see Konow 2003, 2005) that the choices of spectators, in contrast to those of stakeholders, converge, i.e., spectators agree and act on common social norms to a significantly greater degree than stakeholders.

Table 1. Summary Statistics

		First-stage Method	
		Random Division Average (st dev) Slope (st dev)	X Decision Average (st dev) Slope (st dev)
Second Stage Allocator	Stakeholder Y Allocator	\$8.23 (2.45) 0.700 (0.49)	\$6.49 (2.66) 0.933 (0.47)
	Spectator Z Allocator	\$10.13 (0.41) 0.980 (0.68)	\$9.66 (0.99) 1.283 (0.58)

Before turning to an analysis of the hypotheses, we summarize the main results for average second-stage allocations. For this, we run a regression of second-stage allocations averaged across possible first-stage divisions, whereby we suppress the intercept term and include dummies for all four treatments. The parameter estimates on each of the indicator variables provides the statistical differences between that treatment and the three other treatments. We then use the standard errors of these estimates to statistically compare them with each other, which identifies differences from the other treatments. Table 2 summarizes the estimates, t-values and resulting significance levels from those tests. We see that five of the six possible comparisons involving averages indicate significant treatment effects.

Table 2: Between-Treatment Differences of Average Second-Stage Allocations

	DY	DZ	RY
RZ	1.819** (7.46)	0.239 (0.98)	0.950** (3.89)
RY	-0.869** (3.56)	0.711** (2.91)	
DZ	1.581** (6.48)		

** p < .01; * p < .05

4.2 H1: Moral bias in overall social preferences

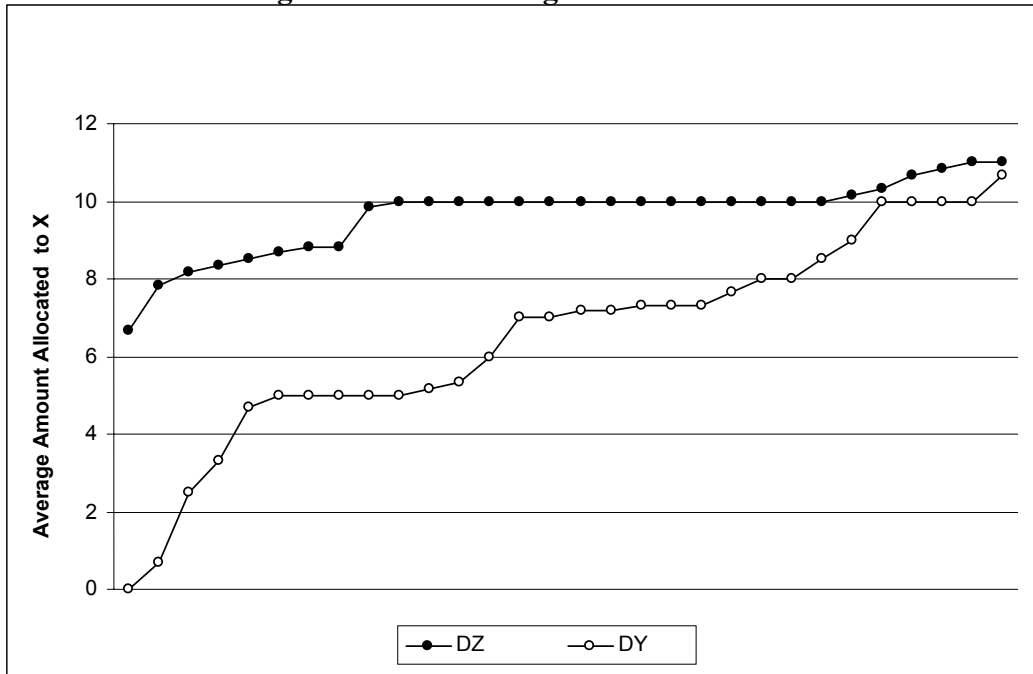
We compare the Decision conditions, where both distributive and reciprocal preferences are relevant, and measure moral bias as differences between the stakeholder Y and spectator Z treatments. Beginning with levels, a *mean moral bias* implies that the allocations of stakeholders to X subjects are significantly less, on average, than those of spectators. We find statistical support for this claim (two-tailed t-test, $p=.0001$). Specifically, stakeholders Y (DY treatment) allocated, on average, \$6.49, or more than 30% less than the \$9.66 allocated by spectators Z (DZ treatment).

The difference in average allocations between stakeholders Y and spectators Z could be due to one of two reasons. One possibility is that most stakeholder Y allocators are behaving differently from most spectator Z allocators, the former giving X subjects less than the latter. Another explanation is that some stakeholder Y allocators are selfish and give little or nothing, whereas others are unselfish and behave like spectator Z allocators. To investigate this, we examine individual decisions by graphing the distribution of the average amounts given to X by stakeholder Y allocators and spectator Z allocators when the initial decision is made by X in Figure 4.

The Y-axis represents the average amounts allocated to X. On the X-axis, second-stage allocators are ordered separately for each treatment from smallest average allocations to

X to largest. As can be seen from the graph, only one stakeholder Y allocator gives zero to her partner. Most stakeholder Y allocators give a positive amount but less than the spectator Z allocators. In fact, 13 of the stakeholders give less to X than the stingiest spectator! In addition, 15 spectators give \$10 on average, or one-half of second-stage stakes, whereas only four of the stakeholders choose that level. Thus, we conclude that the difference between stakeholder Y and spectators Z is not due to a particular mix of selfish and unselfish types. Rather, most stakeholders behave somewhat selfishly.

Figure 4: Second-Stage Allocations to X



Turning now to slopes, a *response moral bias* would mean that the second-stage allocations of stakeholders do not vary as strongly with first-stage divisions as those of spectators. In this case, this implies that the average slope of second-stage allocators is greater in the DZ treatment than in the DY treatment. We do find that spectator Z allocators are significantly more responsive than stakeholder Y allocators to differences in X decisions (two-tailed t-test, $p=.0130$). Spectator Z subjects allocate \$1.28, on average, for every dollar

given to them by X subjects, whereas stakeholder Y subjects allocate only \$0.93, or 27% less than spectators. Comparing these slopes to the slope of 1 associated with inequality aversion, the responsiveness of second-stage stakeholder allocators does not differ from 1 in the DY treatment (two-tailed t-test, $p=.4400$), but, in the DZ treatment, spectators exhibit reciprocal preferences with a slope that is significantly greater than 1 (two-tailed t-test, $p=.0127$). Thus, we find a significant moral bias in overall social preferences, supporting H1.

4.3 H2: Moral bias in distributive preferences

Here we focus on the Random Division treatments, where only distributive preferences are involved, and compare allocations in the RY and RZ treatments. The average allocation is significantly lower in the RY than the RZ treatment (two-tailed t-test, $p=.0001$), indicating the presence of a mean moral bias in distributive preferences. Stakeholders Y allocated, on average, \$8.23, or almost 20% less than the \$10.13 allocated by spectators Z.

Examining slopes, spectators are more responsive than stakeholders: spectators Z allocate, on average, \$0.98 for every dollar given by X subjects, whereas stakeholders Y allocate \$0.70, almost 30% less than spectators Z. Nevertheless, the difference between the two is only marginally significant (two-tailed t-test, $p=.0719$). Comparing these to a slope of 1 shows that the responsiveness of stakeholders is significantly less than 1 (two-tailed t-test, $p=.0024$), whereas that of spectators does not differ significantly from 1 (two-tailed t-test, $p=.8723$). Thus, with regard to distributive preferences, we find a highly significant mean moral bias and a somewhat weaker response moral bias, results that generally support H2.

4.4 H3: Asymmetric reciprocity in stakeholders

To identify stakeholder reciprocity, we compare the DY and RY treatments. Reciprocity is measured by the responsiveness of second-stage allocators to first-stage

divisions. We find that stakeholder Y allocators are over than 30% more responsive to differences in first-stage allocations due to X Decisions than to Random Divisions, although this difference in slopes is only marginally significant (two-tailed t-test, $p=.0648$). If Hypothesis 3 is correct, however, reciprocity could be underestimated in this overall measure, since it combines weak positive reciprocity with strong negative reciprocity. Thus, we examine reward and punishment separately.

Consider the *average* allocations of stakeholder Y allocators facing X decisions versus stakeholder Y allocators facing random divisions (i.e., mean second-stage allocations for DY versus those for RY). Specifically, consider these allocations when the first-stage division to Y is low (\$0, \$2 or \$4) and when the first-stage division to Y is high (\$6, \$8 or \$10). Negative reciprocity should appear as a negative difference between the Decision and the Random treatments when the first-stage division is low; stakeholders allocate less to X when X purposefully transferred a low amount than when the low amount was determined randomly. Positive reciprocity should appear as a positive difference between allocations in the Decision and the Random treatments when the first allocation is high; stakeholders allocate more to X when X purposefully transferred a high amount than when the high amount was determined randomly. Table 3 summarizes second-stage allocations corresponding to Low and High first-stage divisions for each of the treatments. We first focus on the upper row of this table corresponding to Stakeholder Y Allocators.

When stakeholders receive a low amount due to X's actions (as opposed to randomly), they reduce X's payoff by 40% (\$6.19 to \$3.67). This difference is statistically significant, (two-tailed t-test, $p<.001$). In contrast, when stakeholders receive a high amount due to X's actions (as opposed to randomly), they do not reward X. Second-stage stakeholder allocations

in the decision treatment are actually lower by 9% than those in the random treatment (\$9.32 versus \$10.28), although this difference is not statistically significant (two-tailed t-test, $p=.285$). Thus, we find solid support for Hypothesis 3: stakeholders exhibit significant negative reciprocity and insignificant positive reciprocity, i.e., negative reciprocity is stronger than positive reciprocity for stakeholders.

**Table 3: Mean Second-stage Allocations to X (st. dev.)
by Size of First-stage Transfers to Y**

		First-stage Method	
		Random Division Low to Y High to Y	X Decision Low to Y High to Y
Second- stage Allocator	Stakeholder Y Allocator	\$6.19 (2.17) \$10.28 (3.31)	\$3.67 (2.20) \$9.32 (3.55)
	Spectator Z Allocator	\$7.78 (2.51) \$12.49 (2.76)	\$5.92 (1.51) \$13.39 (2.05)

4.5 H4: Asymmetric reciprocity in spectators

Reciprocity in spectators involves comparing the DZ and RZ treatments. Considering first slopes, spectator Z allocations are over 30% more responsive to first-stage divisions in the DZ condition than in the RZ condition. As previously reported, the RZ slope does not differ from 1, and the DZ slope is significantly greater than 1, but the difference between the two slopes is only marginally significant (two-tailed t-test, $p=.0679$). As with stakeholders, therefore, the question arises whether reciprocity is underestimated due to an asymmetry.

The stakeholder asymmetry found in the previous section could be due to moral bias or to a genuine asymmetry in reciprocal preferences found in spectators. We refer again to Table 3 but focus now on the lower row. When Y receives a low amount due to X's actions (as opposed to randomly), spectator Z reduces X's payoff by 24% on average (\$7.78 to \$5.92), (two-tailed t-test, $p=.001$). In contrast, when Y receives a high amount due to X's

actions (as opposed to randomly), spectator Z rewards X by only 7% (\$12.49 to \$13.39), which is not statistically significant (two-tailed t-test, $p=.157$). Thus, spectators also exhibit stronger negative reciprocity than positive reciprocity, corroborating Hypothesis 4.

These results suggest that the asymmetric reciprocity of stakeholders is caused, at least in part, by an asymmetry in the underlying reciprocal preferences. Nevertheless, there could still be a moral bias in reciprocal preferences, if spectators reward more and punish less than stakeholders. This is Hypothesis 5, which we examine in the following subsection.

4.6 **H5: Moral bias in reciprocity**

We analyze asymmetric reciprocity by considering average second-stage allocations separately for Low and High first-stage divisions. We regress these averages on dummies for first-stage method and second-stage allocator and report the results in Table 4. The first (second) column is for second-stage allocations to X related to Low (High) first-stage divisions to Y.

Table 4: Regressions of Average Second-Stage Allocations by Category

	Low	High
	Estimate (t-statistic)	Estimate (t-statistic)
Intercept	5.889** (30.34)	11.369** (41.55)
X Decision	-1.094** (5.63)	-0.014 (0.05)
Stakeholder Y Allocator	-0.961** (4.95)	-1.569** (5.74)
R ² (adjusted)	0.3134	0.2062
N	120	120

** $p < .01$; * $p < .05$; ^ $p < .10$

First, note that the first-stage method (X Decision or Random Division) only matters when the first-stage transfer is low. When X chooses a low transfer to Y, he is punished and

receives significantly less than when a low transfer occurs by random chance. On the other hand, when X transfers a high amount to Y, he is not rewarded, relative to what he receives when that transfer is chosen randomly. This result is consistent with our t-tests above: neither stakeholders Y nor spectators Z significantly reward X.

Second, the stakeholder Y allocator variable is significantly negative when X makes a low transfer, suggesting that stakeholders punish low transfers more than do spectators. In addition, it is significantly negative when X makes a high transfer, suggesting that spectators reward high transfers relative to stakeholders. Thus, there is a moral bias in reciprocity: spectators reward more and punish less than stakeholders (although spectators still punish more than they reward. This evidence supports H5.

5. Conclusion and Discussion

Evidence has been mounting from the laboratory and the field that social preferences are economically relevant and statistically important forces in a variety of settings. They are implicated in involuntary unemployment, strikes and lockouts, product pricing, contract negotiations, and other bargaining behavior. A number of competing motivations have been described and formalized.

With this study we hope to help unify this stream of research. We define and examine moral biases, the impact of self-interest that distorts the expression of social preferences. We find that moral biases exist: stakeholder decisions differ significantly from those of spectators. Specifically, stakeholders exhibit a *mean* moral bias, allocating less, on average, to others than spectators, as well as a *response* moral bias, reacting less than spectators to differences in the allocations of others. These biases are found both with overall social preferences (Hypothesis 1) and with purely distributive preferences (Hypothesis 2). Both stakeholders and

spectators reciprocate but do so asymmetrically: in absolute terms, they punish others for low allocations, but they do not significantly reward them for high ones (Hypotheses 3 and 4). Thus, we conclude that asymmetry in reciprocity is partially driven by an underlying asymmetry in social preferences. However, there is also a moral bias in reciprocity (Hypothesis 5): stakeholders punish low transfers more and reward high transfers less than spectators. These results tend to support Adam Smith's assertion about spectators rewarding kindness but stakeholders not: spectators reward generous X subjects significantly more than stakeholders.

This obscuring effect of self-interest has impaired other attempts to infer social preferences (distributive and reciprocal) and to gauge the magnitude of their force relative to one another and to self-interest itself. Previous studies have come to conflicting conclusions about the importance of these preferences, sometimes suggesting that distributive preferences, reciprocity or both have little or no effect. Our experiment identifies and separates self-interested, distributive and reciprocal preferences. The results reported here allow us to conclude with some confidence that all three forces exert considerable influence on both the level and responsiveness of individual allocation decisions.

In addition, this study helps to clarify the relationships between self-interest, distributive preferences and reciprocity and to quantify them unobscured by moral biases. One surprising result is that stakeholders are less reciprocal, on average, than spectators. This runs counter to the reasonable expectation that stakeholders might respond more strongly to kindness and unkindness directed toward them than would an unimplicated third party on their behalf. Nevertheless, it is consistent with Smith's conjecture regarding spectators and stakeholders.

A related finding concerns the important asymmetry between positive and negative reciprocity that has been seen in previous studies. We find that spectators, like stakeholders, punish more than they reward. However, spectators punish significantly less and reward significantly more than stakeholders, suggesting that the application of reciprocity observed in previous studies included moral biases.

The contrast between spectator and stakeholder reciprocity is potentially helpful in building descriptive models, but it is also a difference that might be important to consider for prescriptive analysis. Social choice models built on the abstract impartial spectator, for example, should consider that the positive reciprocity motive appears to be stronger among spectators than among stakeholders.

We believe that investigation into spectator preferences is an important direction for further research. Such investigations have at least three advantages. First, they allow us to understand and estimate the sometimes intricate social preferences that self-interest might otherwise obscure. This holds the potential for informing theoretical work on how social preferences came to be and how they are instantiated. We think this method can help us determine what behavior is deemed fair or right. Second, by isolating motives apart from self-interest and comparing them with the behavior of stakeholders, this approach enables us to identify the effects of self-interest, the central construct in economics, more precisely. Finally, impartial spectatorship has a long tradition in social choice and moral philosophy, and empirical analysis in this vein might inform normative theory and economic policy.

References

- Abbink, K., Irlenbusch, B., Renner, E., 2000. The moonlighting game - an experimental study on reciprocity and retribution. *Journal of Economic Behavior and Organization* 42, 265-277.
- Akerlof, G.A., Yellen, J., 1990. The fair wage-effort hypothesis and unemployment. *Quarterly Journal of Economics* 105 (2), 255-83.
- Andreoni, J., 1988. Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics* 97, 291-304.
- Andreoni, J., 1993. An experimental test of the public goods crowding-out hypothesis. *American Economic Review* 83 (5), 1317-1327.
- Andreoni, J., Miller, J., 2002. Giving according to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica* 70(2), 737-753.
- Babcock, L., Loewenstein, G., Issacharoff, S., Camerer, C., 1995. Biased judgments of fairness in bargaining. *American Economic Review* 85 (5), 1337-1343.
- Becker, G.S., 1974. A theory of social interactions. *Journal of Political Economy* 82, 1063-1093.
- Ben-Ner, A., Putterman, L., Kong, F., Magan, D., 2004. Reciprocity in a two-part dictator game. *Journal of Economic Behavior and Organization* 53, 333-352.
- Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10 (1), 122-142.
- Blount, S., 1995. When social outcomes aren't fair: the effect of casual attributions on preferences. *Organizational Behavior and Human Decision Processes* 63 (2), 131-144.
- Bolton, G.E., 1991. A comparative model of bargaining: theory and evidence. *American Economic Review* 81(5), 1096-1136.
- Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. *American Economic Review* 90 (1), 166-193.
- Brandts, J., Charness, G., 2000. Hot vs. cold: sequential responses and preference stability in experimental games. *Experimental Economics* 2 (3), 227-238.
- Brandts, J., Solà, C., 2001. Reference points and negative reciprocity in simple sequential games. *Games and Economic Behavior* 36, 138-157.
- Brosig, J., Weimann, J., 2003. The hot vs. cold effect in a simple bargaining experiment. *Experimental Economics* 6 (1), 75-90.
- Carpenter, J.P., Matthews, P.H., 2004. Social reciprocity. Manuscript, Middlebury College.
- Cason, T., and Mui, V-L., 1998. Social influence in the sequential dictator game. *Journal of Mathematical Psychology* 42, 248-265.
- Cason, T., Saijo, T., Yamato, T., 2002. Voluntary participation and spite in public good provision experiments: an international comparison. *Experimental Economics* 5 (2), 133-153.
- Charness, G., 2000. Bargaining efficiency and screening: an experimental investigation. *Journal of Economic Behavior and Organization* 42 (3), 285-304.
- Charness, G., 2004. Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics* 22, 684-708.
- Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117 (3), 817-869.

- Cherry, T.L., Frykblom, P., Shogren, J.F., 2002. Hardnose the dictator. *American Economic Review* 92 (4), 1218-1221.
- Cox, J.C., 2004. How to identify trust and reciprocity. *Games and Economic Behavior* 46, 260-281.
- Cox, J.C., Friedman, D., Gjerstad, S., 2007 A tractable model of reciprocity and fairness. *Games and Economic Behavior* 59, 17-45.
- Croson, R.T.A., 1996. Partners and strangers revisited. *Economic Letters* 53, 25-32.
- Croson, R.T.A., Buchan, N., 1999. Gender and culture: international experimental evidence from trust games. *American Economic Review* 89 (2), 386-391.
- Dickinson, D.L., 2001. The carrot vs. the stick in work team motivation. *Experimental Economics* 4 (1), 107-124.
- Engelmann, D., Strobel, M., 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review* 94 (4), 857-869.
- Falk, A., Fehr, E., Fischbacher, U., 2000. Testing theories of fairness – intentions matter. Working Paper No. 63, Institute of Empirical Research in Economics, University of Zurich.
- Falk, A., Fehr, E., Fischbacher, U., 2003. On the nature of fair behavior. *Economic Inquiry* 41(1), 20-26.
- Falk, A., Fischbacher, U., 2000. A theory of reciprocity. University of Zurich. Accessed at www.iew.unizh.ch/wp/iewwp006.pdf.
- Fehr, E., Fischbacher, U., 2004. Third party punishment and social norms. *Evolution and Human Behavior* 25, 63-87.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114 (3), 817-868.
- Forsythe, R., Horowitz, J.L., Savin, N.E., Sefton, M., 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior* 6, 347-369.
- Gächter, S., Riedl, A., 2005. Moral property rights in bargaining. *Management Science* 51 (2), 249-263.
- Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3, 367-388.
- Hoffman, E., McCabe, K., Shachat, K., Smith, V., 1994. Preferences, property rights and anonymity in bargaining games. *Games and Economic Behavior* 7, 346-380.
- Kachelmeier, S.J., Limberg, S.T., Schadewald, M.S., 1991. A laboratory market examination of the consumer price response to information about producers' costs and profits. *The Accounting Review* 66 (4), 694-717.
- Kahneman, D., Knetsch, J.L., Thaler, R., 1986a. Fairness and the assumptions of economics. *Journal of Business* 59 (4), 285-300.
- Kahneman, D., Knetsch, J.L., Thaler, R., 1986b. Fairness as a constraint on profit seeking: entitlements in the market. *American Economic Review* 76, 728-741.
- Konow, J., 2000. Fair shares: accountability and cognitive dissonance in allocation decisions. *The American Economic Review* 90 (4), 1072-1092.
- Konow, J., 2003. Which is the fairest one of all?: a positive analysis of justice theories. *Journal of Economic Literature* 41(4), 1186-1237.
- Konow, J., 2005. Blind spots: the effects of information and stakes on fairness biases. *Social Justice Research* 18 (4), 349-390.

- Levine, D.K., 1998. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics* 1, 593-622.
- McCabe, K.A., Rigdon, M.L., Smith, V.L., 2003. Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization* 52, 267-275.
- Offerman, T., 2002. Hurting hurts more than helping helps. *European Economic Review* 46, 1423-1437.
- Rabin, M., 1993. Incorporating fairness into game theory and economics. *American Economic Review* 83 (5), 1281-1302.
- Rutström, E.E., Williams, M.B., 2000. Entitlements and fairness: an experimental study of distributive preferences. *Journal of Economic Behavior and Organization* 43, 75-89.
- Smith, A., 1759 [1809]. *The Theory of Moral Sentiments*, Glasgow: R. Chapman.