Double Standards: Social Preferences and Moral Biases

Rachel Croson and James Konow

Loyola Marymount University

March 2007

March 2007

Double Standards: Social Preferences and Moral Biases

Rachel Croson
University of Pennsylvania
Operations & Information Management Department
Philadelphia , PA 19104
crosonr@wharton.upenn.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 social preferences. Among other things, this has led to disagreement about the relative importance of distributive preferences, reciprocal preferences, or both. This paper describes a simple experiment that decomposes behavior into these three forces and examines their interactions without the confounds that have compromised other designs. 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 respond less forcefully to kindness and unkindness towards them than do spectators acting on their behalf. We also find an asymmetry in reciprocity: stakeholders punish but do not reward, whereas spectators both reward and punish. This result suggests that the lack of positive reciprocity found in other studies is not due to an asymmetry in underlying reciprocal preferences but rather to a moral bias by stakeholders in the application of that preference. More generally, we find that all three hypothesized motives have important and significant effects on final allocations.

Keywords: Reciprocity, fairness, justice, moral bias
JEL classification: D63, C91

We wish to thank Alexander Cappelen, Gary Charness, Simon Gächter, David George, Bertil Tungodden, participants at seminars at the Norwegian School of Economics and Business and 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. We assume all responsibility, of course, 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 have increasingly come to embrace the need to enrich traditional models with additional motives. The growing consensus is that integrating such *social preferences* into the analysis can often explain or improve explanations of 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 the often competing forces of self-interest and social preferences in the laboratory and the field has proven to be an important hurdle to inferring the principles that theory needs to incorporate. Most attempts to identify the distinctive impact of social preferences have been by hindered by "moral biases." As used here, moral bias refers to any effect of self-interest that insinuates itself into the observed or reported willingness to act on social preferences.[1] What one often observes under these conditions are "double standards," or the phenomenon that can be defined as "a set of principles that are applied more rigorously to one party than to another, especially oneself." Although the

[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.

existence of moral biases has been well documented, almost no studies of social preferences have disentangled social preferences from self-interest and its distorting effects. This study seeks to isolate experimentally the separate effects of self-interest and two major categories of social preferences (distributive and reciprocal). It is distinct from previous work by establishing a benchmark of pure social preferences and by eliminating any explicit role for strategic behavior, both under various allocation conditions. This permits us to identify each force and the sometimes unexpected interactions among motives.

The Güth et al. (1982) experimental test of the ultimatum game found that many proposers offer nontrivial amounts to responders and that responders often reject small offers, in contradiction to the unique subgame perfect prediction for rational, self-interested agents. This finding framed much of the subsequent literature on social preferences in several important ways: it kindled much of the ensuing empirical and theoretical search for social preference-based explanations, it laid the groundwork for the prominent role that the experimental method has played in investigating such motives, and it set the stage for analyzing social preferences in the context of strategic interaction.

Among possible motives, fairness has been widely suspected (e.g., Bolton 1991, Bolton and Ockenfels 2000, Fehr and Schmidt 1999), although similar "anomalies" in other experiments have been variously attributed to altruism (Becker 1974, Levine 1998), warm glow (Andreoni, 1993), spite (Cason, Saijo and Yamato, 2002), intentions (Rabin, 1993), and trust and reciprocity (Berg, Dickhaut and McCabe, 1995). More recently, though, the central debate has been over whether social preferences are *distributive*, i.e., preferences over outcomes or endstates, or *reciprocal*, i.e., preferences over intentions or player types.[2] The

---

[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

ultimatum design of Blount (1995) provides a nice framework for clarifying this distinction and introducing our design. Suppose a proposer offers a responder $3 out of a sum of $10. Now compare this with the same $3/$7 division, but suppose it had been randomly determined. If the responder's preferences were purely distributive and driven solely by, say, an aversion to inequality, she would either accept or reject this offer, but it would not matter whether it was made by a proposer or determined randomly. If, on the other hand, the responder would reject the proposer's offer as unkind but would accept the randomly determined $3, this is evidence of reciprocity. Specifically, negative reciprocity is manifested when one party punishes another's unkindness, as in this case, and positive reciprocity occurs when one party rewards another's kindness. Nevertheless, a drawback of the ultimatum game is its inability to disentangle the various motives potentially at play.[3]

Another important design in this literature has been the so-called "trust game," introduced by Berg, Dickhaut and McCabe (1995). In this two stage game, most proposers have been found to send some positive amount to responders, who, upon receiving triple the sum sent, usually then return a positive amount of money to the proposer (e.g., see Croson and Buchan 1999). The latter is sometimes taken as a measure of reciprocity and the former of "trust," i.e., behavior based on expectations of reciprocity. The potential problem with these interpretations is that both proposer and responder transfers might be motivated, in part or in whole, by distributive preferences. For this reason, Cox (2004) introduced a "triadic" design that supplements the standard trust game experiment with two treatments that are variations

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.

[3] Falk, Fehr and Fischbacher (2003) introduce a variation on the ultimatum game that produces strong evidence of negative reciprocity: a given unequal offer is more likely to be rejected by a responder if the proposer could have alternately made a more equal offer than if the proposer could only have made a more unequal offer.

on the "dictator game," in which subjects may transfer money to other subjects, but the recipients have no decision.[4] The results show significant evidence of distributive preferences based on the mostly positive transfers of decision makers in dictator-like treatments. Additionally, Cox finds significantly larger transfers by proposers and responders in the trust treatment than among their counterparts in dictator treatments, providing evidence of trust and reciprocity.

The ultimatum game permits only negative reciprocity: the responder can punish a selfish proposer but she cannot, through her choices, vary the reward to a proposer based on the generosity of his offer. Similarly, the trust game focuses on positive reciprocity: the responder can liberally reward proposer generosity but is constrained in punishing proposer selfishness (indeed, a responder is powerless against a proposer who sends nothing). Other designs have allowed expression of both positive and negative reciprocity, including Charness (2004), Offerman (2002) and Abbink, Irlenbusch and Renner (2000), and have found that negative reciprocity appears to be a more powerful motivation than positive reciprocity.

Following on these and other important contributions, recent models have incorporated all three motives, e.g., Charness and Rabin (2002), Cox, Friedman and Gjerstad (2004) and Falk and Fischbacher (2000). The current study has two overarching goals that take this view as its point of departure. First, we decompose the motivational forces behind allocation decisions under various allocation conditions into three parts: self-interest,

---

[4]The 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.

distributive preferences, and reciprocity (including positive and negative reciprocity).[5] Second, we seek individual-level evidence on the interactions between the three motives. The challenge, which we believe our design overcomes, is to decompose these forces and to identify their interactions without the distortions that potentially plague such measures when multiple motives are activated. Whereas previous studies have examined these motives, we believe this is the first to decompose and analyze them in this manner.

Probably the most significant obstacle to these goals is moral biases caused by self-interest, which can have complex and subtle effects on allocative decision-making. One important reason is that self-interest appears not only in unmistakable forms but can also mask itself as social preference when strategic interaction is possible. Consider the ultimatum game. The strategic design of this experiment prompts a self-interested reason for proposer generosity: small offers face an increased risk of rejection by responders such that even purely self-interested proposers have an incentive to make more generous offers, suggesting that ultimatum game offers overstate the true degree of proposer generosity (consistent with the results of Forsythe et al., 1994). Strategic behavior can similarly confound inferences about social preferences in the trust game. Proposers who anticipate sufficient responder reciprocity have a self-interested incentive to transfer money to responders. The inability to read the proposer's motives unambiguously, in turn, potentially confounds inferences about the responder's reciprocal preferences. The responder does not know whether proposer generosity is due to his social preferences, which she would like to reward, or to his self-interest, which she would probably rather punish. It is possible, therefore, that responder transfers understate their true reciprocal preferences.

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

Because of these concerns, we eliminate any explicit reasons for strategic behavior in our experiment. Specifically, 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 either the recipient or another subject allocates $20 between the two subjects from the first stage. We provide the procedural details of the experiment in the following section of the paper, but one important treatment variable concerns the identity of the second stage allocator. In one version this is one of the two subjects from the first stage, whereas in the other version it is a third party, who is paid a fixed fee unrelated to the division of $20 she chooses between the two first stage subjects. In this way, one can distinguish the decisions of an implicated "stakeholder" in a second-person relationship from those of an impartial "spectator" who is expressing pure preferences over the allocations of other persons. This permits us to address a number of questions about social preferences, e.g., the conjecture, implicit in the Adam Smith quote at the start, that reciprocity manifests itself less strongly in a stakeholder than in a spectator.

Another issue this design allows us to address concerns the previously-cited asymmetry between positive and negative reciprocity among stakeholders. One possibility is that this asymmetry is a genuine characteristic of the underlying social preference, i.e., even spectators would reward less strongly than they punish. An alternative hypothesis is that reciprocity considered alone prescribes symmetric reward and punishment but that adding self-interest produces an asymmetric moral bias: stakeholders are more willing to sacrifice self-interest to punish than to reward. In contrast to previous studies, this experiment can distinguish between these hypotheses.

Third party preferences have been found in previous work. Indeed, many of the stories underlying evolutionary economics that support social preferences rely on these types of motives (e.g. Frank 1988, Boyd et al. 2003, Güth and Ockenfels 2000). 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 and differ, therefore, in several ways. For one thing, their third parties are stakeholders rather than spectators, i.e., they must incur a cost to punish.[6] This underscores the strength of this preference, but it also opens the possibility of a moral bias in the measurement of this preference.[7] Second, they focus on negative reciprocity, whereas we wish to 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 never informed of the second stage in order to minimize or eliminate strategic considerations. This design feature is lacking in most previous studies.

---

[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. Charness and Rabin (2002) do include genuine third party spectators in two of the many games they report. These, however, involve only two sets of comparisons in binary choice dictator games that focus on preferences for surplus maximization and that do not explore distributive and reciprocal preferences over a wide range of allocation conditions, as we set out to do in the current study.

[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 dismisses social preferences altogether, third party decisions should have salience. Moreover, if they did not, one would probably expect either no or a very simple pattern of decisions, but, as later reported in this paper, third parties demonstrate considerable sophistication and converge in their decisions significantly more than stakeholders.

Our analysis focuses on the effects of each of the three motives on levels of second stage allocator transfers as well as on their responsiveness at an individual level to changes in first stage allocations. The design takes account not only of the distortionary effect of self-interest but also attempts to minimize any contextual cues that might activate more complex distributive preferences and, thereby, introduce confounds. For example, Andreoni and Miller and Charness and Rabin have found that, when total stakes are variable, many subjects exhibit an "efficiency motive," i.e., a desire to maximize surplus, even at a personal cost. Subjects have also been observed to respond to a sense of desert when they earn their allocations through a task, as in Konow (2000), Cherry, Frykblom and Shogren (2002), and Rutström and Williams (2000), or are led to believe they have earned their allocations, as in Hoffman, et al. (1994). For these reasons, we use fixed stakes in both stages that are endowed, rather than earned. Moreover, the fact that the second round stakes are double those in the first round gives considerable latitude given to the expression of both positive and negative reciprocity.

The paper is organized as follows. Section 2 describes the experimental design and procedures, section 3 presents and analyzes the results of the experiment, and section 4 contains the concluding remarks.

2. Description of the Experiment

2.1 Experimental Design

The experiment is a two stage dictator game. In the first stage, a sum of $10 is distributed between each of two paired subjects in groups denoted X and Y. One treatment variable has to do with the method of this distribution. In the "Group X Decision" condition, each subject in the X group receives $10, which he can divide, as dictator, between himself and an anonymous counterpart in group Y. Each X subject can divide the sum in any even dollar amounts, i.e., there are six possible (x,y) divisions in $2 increments from (10, 0) to (0, 10). 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 (as in Blount, 1995).

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 “Group Y Allocator” condition, each Y subject chooses how much to allocate to herself and her X counterpart from the first stage in any one-dollar increment. The “Group Z Allocator” condition is 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 both of these treatments: Y (or 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.

Figure 1. Experimental Design

		First-stage Method	
		Random Division	Group X Decision
Second Stage Allocator	Group Y Allocator	RY (X, **Y**)	DY (**X**, **Y**)
	Group Z Allocator	RZ (X, Y, **Z**)	DZ (**X**, Y, **Z**)

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

[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.

These conditions are crossed to yield a 2 First-stage Method (Group X Decision, Random Division) × 2 Second-stage Allocator (Group Y Allocator, Group Z Allocator) factorial design. Thirty pairs or triples, respectively, participated in each cell in an across-subjects design. This is summarized in Figure 1, which also presents the abbreviations for these treatments that we will use.

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, particularly 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 possible positive second-stage reciprocity (i.e., rewards) in response to Group 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 if first-stage divisions had been in any dollar amount.

[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 such differences, although they are typically relatively small and will not, therefore, be the focus of later data analysis.

Of the four treatments, treatment DY is closest in spirit to the standard trust game.[11] One important difference is the 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 transfers 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 subject 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 distortion from strategic interests.

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.

[11]Our treatment, however, involves fixed stakes since no transfers are ever multiplied. In this respect, ours is also similar to the Specific Reciprocity treatment of Ben-Ner, et al. (2004). Some differences with that study include the facts that their experiment lasted about 2 hours, since the subjects also completed a lengthy survey for which they were paid a separate fixed $15, second-stage allocators made only one choice about what to send after finding out what their first-stage counterpart had sent, and 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 two phases in Konow (2000), and to designs used in other social preference experiments.

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 allocators, therefore, have no basis for expressing reciprocity, since X subjects are unable by design to indicate 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 (fairness). 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	Group X Decision
Second Stage Allocator	Group Y Allocator	RY Self Interest, Distributive	DY Self Interest, Distributive, Reciprocal
	Group Z Allocator	RZ Distributive	DZ Distributive, Reciprocal

For example, two comparisons reveal the aforementioned moral biases, i.e., the effects of self-interest on the expression of social preferences. In the RY and RZ treatments, initial allocations are random, and second-stage allocators can act on their distributive, but not

[13] One might wonder why we test for distributive preferences in the RZ treatment instead of simply assuming, as we did earlier for purposes of illustration, that impartial distributive preferences reduce to equal splits of the total amount from both stages. 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 assuming any specific form.

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 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 Y self-interest (versus 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 Z subjects with no personal stakes, i.e., spectators. 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 subjects) rather than spectators (Z subjects), 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.

2.2 Experimental Procedures

After registering, receiving their show-up fee and being seated (randomly, in the case of X/Y sessions), subjects receive a form with the first-stage instructions and allocation information. The experimenter then reads the instructions aloud. The experiment is run on paper and is conducted single-blind: each subject is identified only by a subject ID, which only he and the experimenter know. Subjects are told that they will never know the identity of their counterparts. All subjects are provided with the same information, but no one is informed, at this point, of the second stage. In the Group X Decision condition, X subjects choose one of the six X/Y divisions by circling it. In the Random Division condition, one of the lines is already circled on the X forms only.[14] These forms are then collected.

Now subjects are told for the first time of a second stage of the experiment and are informed that this is the final decision. They receive forms with the second-stage instructions and space, if applicable, for decisions. In RY and RZ sessions, subjects are informed that the $10 stakes in the first stage were randomly divided between X and Y subjects. In the DY and DZ sessions, they are told that X made this decision. Then, the second-stage allocator (subject Y in the RY and DY sessions and subject Z in the RZ and DZ sessions) chooses how to allocate the $20 between X and Y for each of the six possible first-stage divisions. These forms are then collected, and, while payments are being calculated, all subjects complete a questionnaire that asks demographic information. Subjects in decision-making roles are also asked why they chose as they did as well as a series of hypothetical questions about how they would have chosen if the first-stage divisions were made under the other condition or if they

[14] The actual random division of payoffs uses the full range of possible (x, y) divisions, {(10, 0), (8, 2), …, (0, 10)}. It is from a symmetric distribution that does not, therefore, favor X or Y, and that gives a higher probability to more equal divisions, as is common with actual X decisions. Specifically, the frequency of an allocation with $x in these treatments is $f = .5 \cdot (19 - 3 \cdot |x - 5|)$, which produces a simple, piecewise linear distribution with these attributes.

had been in one of the other decision-making roles. Subjects turn in these forms, sign for payments and are free to depart.

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 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, 96% of the 293 subjects who responded to the question, indicated that they would be willing to participate in an economics experiment again.

3. Results and Analysis

3.1 Summary of Results

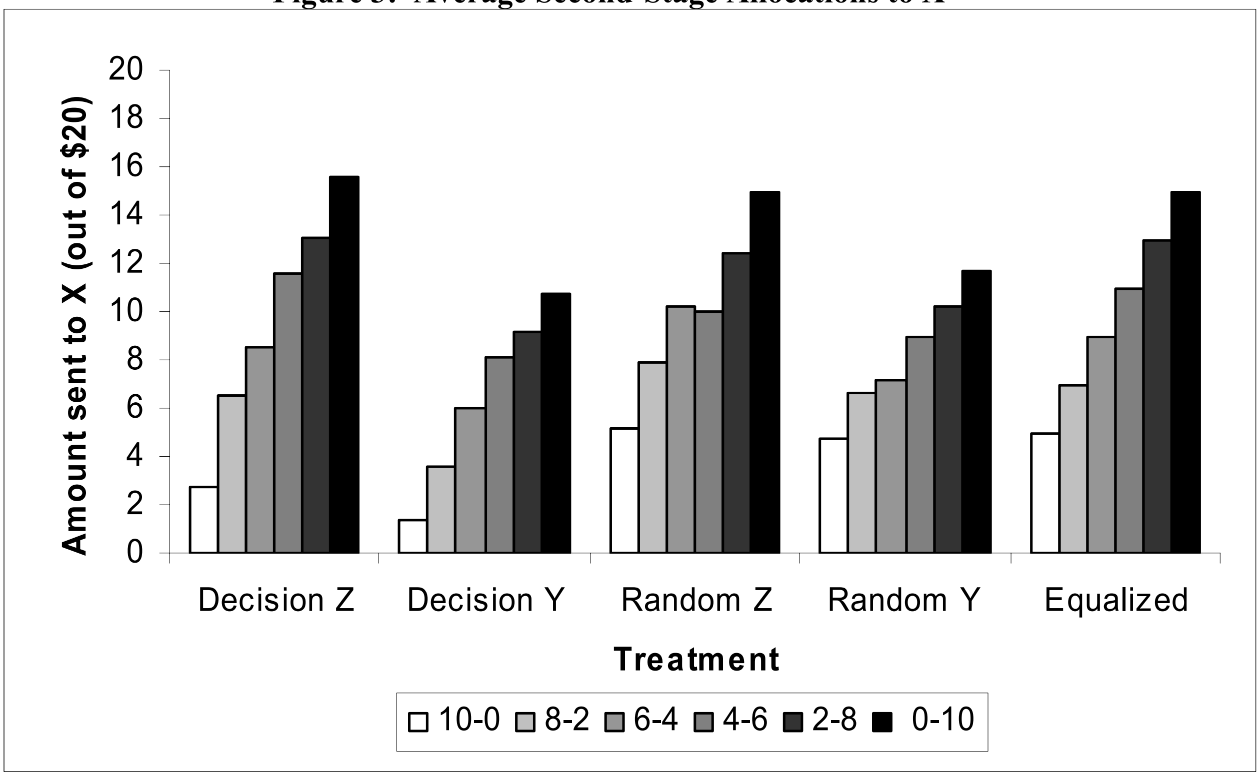
Our primary interest is in the allocations made in the second stage of decision-making. We note the average first-stage transfer to Y in the Group X Decision sessions was $3.07 (st. dev. 1.92), or around 30% of the available amount. These transfers are comparable to those in other dictator studies.

The decisions of the second-stage allocators are our primary interest and were made using the strategy method, i.e., for each possible stage one division without knowledge of the actual outcome of that phase. Figure 3 displays these second-stage allocations, where each set

[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).

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 stage one divisions. It is apparent from the increasing bar heights that, as Y's first-stage allocation increases, the amount allocated to X in the second stage increases (with a single minor exception at 4-6 in the RZ treatment). For comparison, the final set of bars denoted "Equalized" describes the amounts that would have to be sent to X in order to equalize earnings.

Figure 3: Average Second-Stage Allocations to X



We compare second-stage allocations across treatments in three 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. These can be related to the first-stage method (Group X Decision or Random Division) and to the second-stage allocator (Y or Z). 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 (Y or Z) allocated to X for each possible division of the $10 in the first stage. This involves calculating the *slope* of the amounts allocated to X in stage two with respect to the amounts allocated to Y in the stage one. This can be interpreted as the slope of the bars in each treatment in Figure 3.[16] Finally, we decompose the types of preferences as outlined in Figure 2 above. Note that for each of these measures, we have only one observation per subject in our analysis.

Table 1 presents summary statistics of the second-stage allocations for the four treatments. Beginning with levels, we compare across treatments the second-stage allocations to X averaged over all possible first-stage divisions. We find that Group Y allocators allocate significantly less to Group X than do Group Z allocators: the average allocation is significantly lower in the Random Y than the Random Z treatment (t=4.198, p=.0001) and in the Decision Y than the Decision Z treatment (t=5.924, p=.0001). Group Y allocations are also significantly greater in the Random Division treatment than in the Group X Decision treatment (t=2.516, p=.0147) as are Group Z allocations (t=2.440, p=.0178). Note also that, when Y is the second-stage allocator, these averages are significantly less than the $10 that would be needed to equalize totals, both in the Random Division treatment (t=3.94, p=.0005) as well as in the Group X Decision treatment (t=7.199, p<.0001). The average amounts allocated to X by Group Z allocators, on the other hand, do not differ at conventional levels of significance from $10, in either the Random Division treatment (t=1.77, p=.0965) or the Group X Decision treatment (t=1.90, p=.0667). Another difference between Y and Z

[16]The 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, then average over the individuals.

allocators is that the standard deviations are significantly greater in the Y treatments than in the Z treatments (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 (here Group Z allocators), in contrast to those of stakeholders (here Group Y allocators), are unbiased and 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)**	Group X Decision Average (st dev) **Slope (st dev)**
Second Stage Allocator	Group Y Allocator	$8.23 (2.45) **0.700 (0.49)**	$6.49 (2.66) **0.933 (0.47)**
	Group Z Allocator	$10.13 (0.41) **0.980 (0.68)**	$9.66 (0.99) **1.283 (0.58)**

Turning now to slopes, we examine the responsiveness of second-stage allocators, averaged across all possible first-stage divisions, for each treatment. We find support for Adam Smith's suggestion at the start of this paper that the appreciation of the stakeholder (Group Y allocator) for kindness (Group X decisions) is less than that of the impartial spectator (Group Z allocator): Group Z allocators are significantly more responsive than Group Y allocators to differences in Group X decisions (t=2.564, p=.0130). Other differences in slopes across treatments, however, are only marginally significant.[17] Comparing these slopes to the slope of 1 associated with equalization of cumulative earnings, the

[17] Specifically, Group Z allocators are marginally more responsive than Group Y allocators in the Random Division treatments (t=1.834, p=.0719), and second-stage allocators respond marginally more strongly to differences in first-stage allocations due to Group X Decisions than to Random Divisions, both in the case of Group Y allocators (t=1.882, p=.0648) and of Group Z allocators (t=1.860, p=.0679).

responsiveness of second-stage allocators does not differ from 1 in the RZ (t=.1622, p=.8723) and DY treatments (t=.7830, p=.4400). The responsiveness, however, of Group Y allocators in the Random Division treatment is significantly less than 1 (t=3.33, p=.0024), and that of Group Z allocators in the Group X Decision treatment, on the other hand, is significantly greater than that suggested by inequality aversion alone (t=2.656, p=.0127).

3.2 Regression Analysis of Average Second-Stage Allocations

Here we apply regression analysis to identify the impact of the first-stage method and of the second-stage allocator on average second-stage allocations. First, we calculate the amount, averaged across the six possible first-stage divisions, that each second-stage allocator allocated to X. Thus, we have one observation per subject. These amounts are regressed on a dummy variable for the first-stage method (1=Group X Decision) and a dummy variable for the second-stage allocator (1=Group Y Allocator). Regression results are reported in Table 2.

Table 2: Regression of Average Second-Stage Allocations

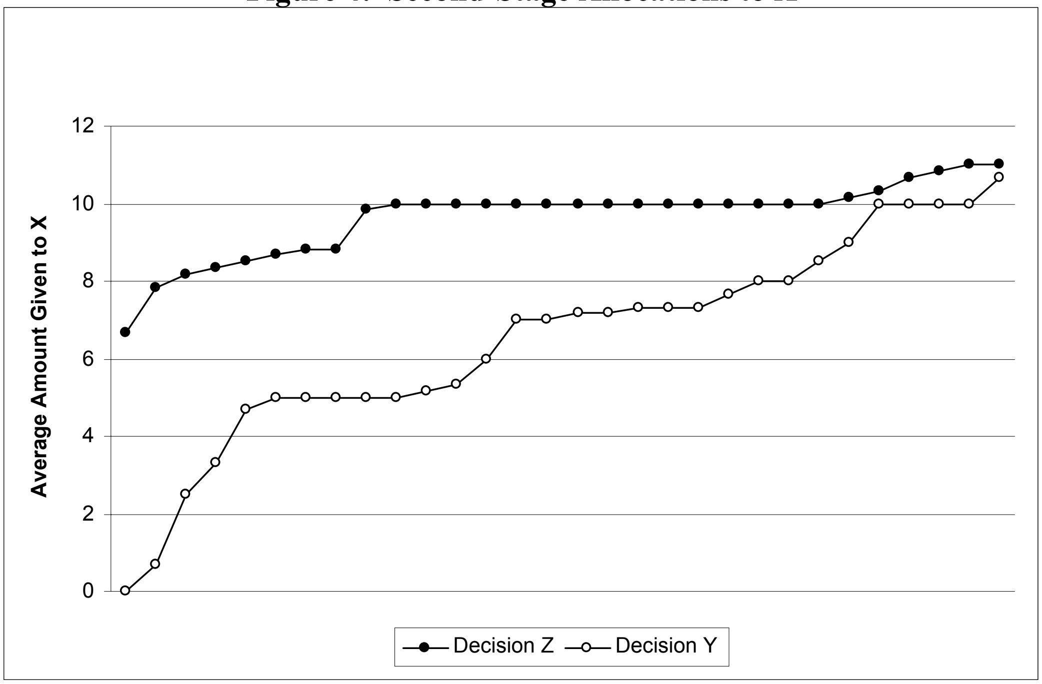
	Estimate (t-statistic)
Intercept	8.629** (49.56)
Group X Decision	-0.554** (3.18)
Group Y Allocator	-1.265** (7.27)
R^2 (adjusted)	0.3387
N	120

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

There are two significant effects. When X decides the first-stage allocation, he receives significantly less than when the first division is random. As will be shown below, this is primarily due to negative reciprocity, whereby first-movers are punished for low offers with greater frequency and force than they are rewarded for high offers. Second, when Y makes the second allocation, X receives significantly less than when Z makes it. This can be attributed to

Y's self-interest: she keeps a larger share for herself, on average, whereas Z treats X and Y more equally.[18]

Figure 4: Second-Stage Allocations to X



The difference in average allocations between stakeholders (Y) and spectators (Z) could be due to one of two reasons. One possibility is that most Y allocators are behaving differently from most Z allocators, viz., the former are giving X subjects less than the latter. Another explanation is that some Y allocators are selfish and give little or nothing, whereas others are unselfish and behave like Z allocators. To investigate this, we examine individual decisions by graphing the distribution of the average amounts given to X by Y allocators and Z allocators when the initial decision is made by X in Figure 4. As can be seen from the graph, only one Y allocator gives zero to his partner. Most Y allocators give a positive

[18] We also ran this and other regressions with interaction terms. Here the coefficient on the Group X Decision/Group Y Allocator term was negative but significant only at the 10% level and the significance of the other independent variables remains. Interaction terms in the other regressions were insignificant even at the 10% level, so these results are not reported. Coefficients and significance levels for the variables of interest are thus robust to the inclusion of this interaction term. In addition, we run a probit regession which accounts for the truncated dependent variable (amounts sent are bounded by 0 and $20, while slopes are bounded by 0 and 1). Again, the results are robust to this alternative specification. These regressions are available from the authors upon request.

amount but less than the Z allocators. In fact, 13 of the Y allocators give less to X than the stingiest Z allocator! In addition, 15 Z allocators give $10 on average, or one-half of second stage stakes, whereas only four of the Y allocators choose that level. Thus, we conclude that the differences between stakeholders Y and spectators Z is not due to a mix of selfish and unselfish types. Rather, most stakeholders behave somewhat selfishly.

3.3 Regression Analysis of Responsiveness of Second-Stage Allocations

The second part of the regression analysis identifies the responsiveness, or slope, of the second-stage allocation to the first-stage method and the second-stage allocator. We calculate these slopes at the individual level. Each second-stage allocator indicated how much she would allocate to X under each of six possible first-stage divisions. This produces five slopes, or changes in allocations to X for each of the five differences between the first-stage divisions. The responsiveness of second-stage allocations is measured as an average, viz., the average of these slopes (again, based on only one observation per participant). These are then regressed on the first-stage method (1=Group X Decision) and the second-stage allocator (1=Group Y Allocator). Regression results are reported in Table 3.

Table 3: Regression of Second-Stage Slopes

	Estimate (t-statistic)
Intercept	0.974** (19.10)
Group X Decision	0.134** (2.63)
Group Y Allocator	-0.158** (3.09)
R^2 (adjusted)	0.1082
N	120

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

There are two significant effects. In particular, the second-stage allocator is significantly more sensitive to differences in first-stage allocations when they are made by X than when they are randomly determined. That is, both Y and Z allocators reciprocate when the first allocation is due to X's choice. The second main effect is that Y's allocations are significantly *less* sensitive to X's initial allocation than are Z's allocations. As discussed below, this is primarily due to spectators Z both rewarding and punishing X outcomes, whereas stakeholders Y punish but do not reward. The lack of a reward, therefore, makes Y less responsive on average to X's actions than Z.

3.4 Decomposing Preferences

As we saw in Figure 2, the four treatments can be compared to isolate the effects of different motives. The baseline RZ treatment (Random division, Group Z allocator) can be used to gauge purely distributive preferences. Comparing that treatment with RY, when the first division is made randomly but the stakeholder makes the second allocation, suggests the scope of self-interest in the absence of reciprocity. A comparison of treatments DY and DZ, when X makes the first decision, measures self-interest in the presence of reciprocity. Finally, we can estimate the effects of reciprocal motives by comparing treatments RZ and DZ (in the absence of self-interest) and RY and DY (in the presence of self-interest).

To test for differences in levels, we run a regression of second-stage allocations averaged across possible first-stage divisions. In this regression, however, 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 4, below, summarizes the estimates, t-values and resulting significance levels from those tests.

Table 4: Between-Treatment Differences of Average Second-Stage Allocations

	Decision Y	Decision Z	Random Y
Random Z	1.819** (7.46)	0.239 (0.98)	0.950** (3.89)
Random Y	-0.869** (3.56)	0.711** (2.91)	
Decision Z	1.581** (6.48)		

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

To test for differences in responsiveness, we conduct pair-wise comparisons of slopes across treatments, similar to those reported in Table 4. Dependent variables are each individual's average slope and independent variables are dummies for all four treatments (again, suppressing the intercept term). Table 5, below, summarizes the estimates, t-values and resulting significance levels from those comparisons.

Table 5: Between-Treatment Differences of Second-Stage Slopes

	Decision Y	Decision Z	Random Y
Random Z	0.023 (0.32)	0.152* (2.09)	0.140 (1.93)
Random Y	0.117 (1.61)	0.292** (4.03)	
Decision Z	0.175* (2.41)		

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

It appears above that treatment effects are more widespread for levels than slopes. Below we summarize the magnitude of differences in the levels and slopes of second-stage transfers due to self-interest (both with and without reciprocity) and to reciprocity (both with and without self-interest). The differences are significant at the 5% level, except where otherwise stated.

Self-interest with distributive preferences

In the Random Division condition, stakeholders Y (RY treatment) allocated, on average, $8.23, or almost 20% less than the $10.13 allocated by spectators Z (RZ treatment). Thus, self-interest influences allocations in the presence of distributive preferences by reducing the amounts given. Stakeholders were less responsive than spectators (the slope for the former is about 30% less than that for the latter), although this difference is not significant at conventional levels, consistent with the absence of any role for reciprocity in these treatments.

Self-interest with distributive preferences and reciprocity

In the Group X Decision condition, stakeholders Y (DY treatment) allocated, on average, $6.49, or more than 30% less than the $9.66 allocated by spectators (DZ treatment). Stakeholders were also almost 30% less responsive than spectators to differences in first-stage allocations (0.933 versus 1.283). Self-interest, therefore, is a significant force in the presence of both distributive and reciprocal preferences lowering amounts sent and making participants less reciprocal..

Reciprocity with distributive preferences

When the second-stage allocators were spectators, they allocated, on average, amounts insignificantly different from $10, viz., $9.66 in the Group X Decision (DZ) treatment and $10.13 in the Random Division (RZ) treatment. They were, however, about 30% more responsive when X decided the first-stage division (slope of 1.283 in DZ) than when these amounts were random (slope of 0.980 in RZ). The DZ slope coefficient means that when X subjects send their Y counterparts an additional $1, Z subjects reward X subjects with about

30 cents more than the amount dictated by distributive preferences alone. Thus we observe a significant effect of reciprocity when self-interest is involved.

Reciprocity with self-interest and distributive preferences

When both self-interest and distributive preferences are relevant, the effect of reciprocity is to lower average transfers. Second-stage allocators who were stakeholders transferred, on average, more than 20% less to their X counterparts when the first-stage division was by Group X Decision ($6.49 in DY) than by Random Division ($8.23 in RY). The slope in the DY treatment is about 30% greater than that in the RY treatment, but this difference is not significant at conventional levels. We will see in the next section that stakeholders do, in fact, reciprocate significantly, but that this effect is tempered by the fact that they do so asymmetrically, i.e., they differ with respect to negative versus positive reciprocity. This asymmetry (and its absence in the spectator allocations) can be traced to the differential effects of self-interest on reciprocity, further discussed in the next subsection.

3.5 Positive versus Negative Reciprocity

In this section we further explore this last result by considering the impact of the identity of the second-stage allocator separately for reward and punishment. Rather than impose cutoff points for first-stage dictator behavior that is deemed worthy of reward or punishment by assumption, we let this be guided by the data. Specifically, we examine the *average* responses of impartial Z allocators to X decisions versus their responses to Random divisions (i.e., the mean second-stage allocation for DZ minus that for RZ) for each first-stage division. We find that this difference is negative when the first-stage allocations to Y equal $0, $2 or $4, which we will call Low, indicating negative reciprocity. This difference is positive, however, when first-stage allocations to Y equal $6, $8 and $10, which we will call

High, indicating positive reciprocity. Table 6 summarizes statistics for second-stage allocations corresponding to Low and High first-stage divisions for each of the treatments.

Table 6: Mean Second-stage Allocations to X (st. dev.) by Size of First-stage Allocations to Y

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

Table 7 reports the results of two separate regressions. In the first column, the dependent variable is the average amount allocated to X in the second stage for the case in which X allocates a low amount to Y in the first stage. In the second column, it is the average amount allocated to X in the second stage for the case in which X allocates a high amount to Y in the first stage.

Table 7: Regressions of Average Second-Stage Allocations by Category

	Low Estimate (t-statistic)	High Estimate (t-statistic)
Intercept	5.889** (30.34)	11.369** (41.55)
Group X Decision	-1.094** (5.63)	-0.014 (0.05)
Group 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 (Group X Decision or Random Division) only matters when the first-stage allocation is low. When X chooses a low allocation to Y, he is punished and receives significantly less than when a low allocation occurs by random chance. On the other hand, this analysis provides no evidence of reward. Actually, an examination of Z allocations does, as previously discussed, reveal a tendency on the part of spectators to reward X generosity, but as we will see below, this is crowded out by the lack of reward exhibited by interested Y subjects. Second, consider the Y allocator variable. It is significantly negative when X makes a low allocation, suggesting that stakeholders punish low allocations more than do spectators. In fact, it is significantly negative even when X makes a high allocation, suggesting that stakeholders fail to reward high allocations compared to spectators. Thus, spectators reward more and punish less than stakeholders.[19]

This last fact can also be seen by comparing transfers of second-stage allocators in decision treatments to those in random division treatments. Table 8 reports the percentages of Z and Y transfers that punished X for low transfers to Y or that rewarded X for high transfers to Y. Punishment refers to the percentage of Z (or Y) allocations in the X Decision treatment that are less than the mean Z (or Y) allocation in the RZ (or RY) treatment, calculated separately for each of the three low X transfers. Reward is similarly defined for high X transfers as the percentage of Z (or Y) allocations in the X Decision treatment that are greater than the mean Z (or Y) allocations in the RZ (or RY) treatment.

[19]One astute reader wondered about the within-subject correlation of reciprocity and fairness. To investigate this, we restrict our attention to the Decision Y treatment, where Y allocators have the opportunity to be reciprocal. We correlate the slope measure that we use for reciprocity of these agents with the average allocations those same subjects reported they would have made if they had been in the Random Y treatment in the post-experimental questionnaire. This correlation is quite low (.067) and insignificant. Thus, it does not seem to be the case that subjects who care more about fairness are also more reciprocal.

Table 8: Percent of Allocations Punishing and Rewarding in Decision Treatments

	Decision Z	Decision Y
Low/Punishment	100%	90%
High/Reward	97%	60%

Almost all second-stage allocators engaged in punishment (100% of Z's and 90% of Y's), and there is no significant difference between the proportion of second-stage allocators who punished. Even more unexpected and important, however, are the comparisons among rewards. Ninety-seven percent of disinterested spectators (Z players) rewarded generous offers by X subjects compared with only 60% of stakeholders (Y players). A t-test of proportions shows this difference to be a significant ($t=2.86$, $p=.006$). Thus, when X has made a generous decision and Y has an opportunity to respond, 40% of the time he is not rewarded and instead X leaves with less money than he would have had the (same) first-stage division been made randomly.

3.6 Summary

The preceding analysis produces a number of conclusions. First, distributive preferences here do not differ greatly from inequality aversion based on spectator allocations that come close to equalizing payoffs to subjects from both stages. Studies that contain more contextual elements, such as variable stakes, earned endowments or information about gender, age or needs, often find significant deviations from equality. This study, however, like many others, deliberately omits such elements, in this case, in order to simplify the distributive task and, in that way, to gauge better its impact relative to other forces. Second, implicated stakeholders are less generous than disinterested spectators: self-interest reduces average second-stage allocations by 20-30% and causes second-stage allocators to be about 30% less responsive to first-stage allocations. Third, both spectators and stakeholders are about 30% more responsive to differences in first-stage divisions when X chooses them than when they are random, suggesting a role for reciprocity. Furthermore, our results tend to substantiate Adam Smith's assertion that the stakeholder's gratitude will not always correspond to a kindness, whereas the spectator's will: spectators allocate, on average, $1.28 to X subjects for every $1 the latter send to Y subjects, whereas Y stakeholders barely return the $1. This is because stakeholders are generally less generous and less responsive than spectators in both the random division and in the Group X decision treatments. Further investigation reveals an asymmetry in punishment and reward behavior among spectators and stakeholders. When subject X transfers are low, almost all spectators and stakeholders punish them with lower second-stage allocations than in the random division cases. But when subject X transfers are high, spectators and stakeholders behave differently: significantly more spectators reward X generosity than do interested stakeholders.

4. 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 discovered and formalized.

With this study, we hope to clarify and help unify this stream of research. Moral biases are known to distort the expression of social preferences and to contribute to double standards in their application. This obscuring effect of self-interest has impaired attempts to infer the general principles of social preferences and to gauge the magnitude of their force relative both to one another and to self-interest itself. We employ a new experiment to identify and separate self-interested, distributive and reciprocal preferences without the confounds that are potentially present in existing games. One benefit of this is to establish more accurate measures of these three motives. Previous studies have come to conflicting conclusions about their importance, sometimes suggesting that distributive preferences, reciprocity or both have little or no effect. The results reported here allow one 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 interactions between self-interest, distributive preferences and reciprocity and to quantify these unobscured by moral biases. In this category, one surprising result that emerged is that stakeholders are less reciprocal 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 acting purely on social preferences engage in both reward and punishment, suggesting that the asymmetric responses of stakeholders are not due to an asymmetry in the underlying preference. Instead, the source appears to be a moral bias in the application of reciprocity: stakeholders do not reward but they do punish, indeed, more than spectators. In fact, the results of Dickinson (2001) and Andreoni, Harbaugh and Vesterlund (2003) might be seen as indicating that such behavior is consistent with more efficient incentive systems. They find that agents are more likely to respond to negative than to positive incentives. The contrast between spectator and stakeholder reciprocity is potentially helpful is building descriptive models. But it is a difference that might also be important to consider for prescriptive analysis. Social choice models built on the impartial spectator model, for example, should consider the positive reciprocity motive, even if it does not figure prominently 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 one to understand and estimate better the sometimes intricate and interacting social preferences, which self-interest might otherwise obscure. This holds the potential for informing theoretical work on how social preferences came to be and how they are currently instantiated. We think this method can help us determine what behavior is deemed fair or right. Second, by isolating the effects of motives apart from self-interest and comparing them with behavior of stakeholders, this approach enables one to identify more precisely the effects of self-interest itself, the central construct in economics. Finally, impartial spectatorship has a long and honored tradition in social choice and moral philosophy. This empirical agenda, therefore, has implications for normative studies, including theoretical work as well as policy research.

References

Abbink, Klaus, Irlenbusch, Bernd, and Renner, Elke, 2000, "The Moonlighting Game - An Experimental Study on Reciprocity and Retribution," *Journal of Economic Behavior and Organization*, 42, 265-277.

Akerlof, George A. and Janet Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *Quarterly Journal of Economics* 105:2, pp. 255-83.

Andreoni, James, 1988, "Why Free Ride? Strategies and Learning in Public Goods Experiments." *Journal of Public Economics*, *97*, 291-304.

Andreoni, James, 1993, "An Experimental Test of the Public Goods Crowding-Out Hypothesis," *American Economic Review*, *83*(5), 1317-1327.

Andreoni, James, Harbaugh, William, and Vesterlund, Lise, 2003, "The Carrot or the Stick: Rewards, Punishments, and Cooperation," *American Economic Review*, *93*(3), 893-902.

Babcock, Linda, Loewenstein, George, Issacharoff, Samuel and Camerer, Colin, 1995, "Biased Judgments of Fairness in Bargaining," *American Economic Review*, *85*(5), 1337-1343.

Becker, Gary S., 1974, "A Theory of Social Interactions," *Journal of Political Economy*, *82*, 1063-1093.

Ben-Ner, Avner, Putterman, Louis, Kong, Fanmin and Magan, Dan, 2004, "Reciprocity in a Two-Part Dictator Game," *Journal of Economic Behavior and Organization*, *53*, 333-352.

Berg, Joyce, Dickhaut, John and McCabe, Kevin, 1995, "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, *10*(1), 122-142.

Blount, Sally, 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, Gary E., 1991, "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, *81*(5), 1096-1136.

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

Boyd, Robert, Gintis, Herbert, Bowles, Samuel and Richerson, P.J. 2003, "The Evolution of Altruistic Preferences," *Proceedings of the National Academy of Sciences*, 100, 227-238.

Brandts, Jordi, and Charness, Gary. "Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games," *Experimental Economics*, *2*:3, 2000, 227-238.

Brosig, Jeannette, and Weimann, Joachim, 2003, "The Hot vs. Cold Effect in a Simple Bargaining Experiment," *Experimental Economics*, *6*(1), 75-90.

Carpenter, Jeffrey Paul and Matthews, Peter Hans, 2004, "Social Reciprocity," manuscript, Middlebury College.

Cason, Timothy, and Mui, Vai-Lam, 1998, "Social Influence in the Sequential Dictator Game," *Journal of Mathematical Psychology*, *42*, 248-265.

Cason, Timothy, Saijo, Tatsuyoshi and Yamato, Takehiko, 2002, "Voluntary Participation and Spite in Public Good Provision Experiments: An International Comparison," *Experimental Economics*, *5*(2), 133-153.

Charness, Gary, 2004, "Attribution and Reciprocity in an Experimental Labor Market." *Journal of Labor Economics 22*, 684-708.

Charness, Gary, and Rabin, Matthew. 2002, "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117:3, 817-869.
Cherry, Todd L., Frykblom, Peter, and Shogren, Jason F., 2002, "Hardnose the Dictator," *American Economic Review*, *92*(4), 1218-1221.
Cox, James C., Friedman, Daniel, and Gjerstad, Steven, 2004, "A Tractable Model of Reciprocity and Fairness," University of Arizona; University of California Santa Cruz, May.
Cox, James C., 2004, "How to Identify Trust and Reciprocity," *Games and Economic Behavior*, *46*, 260-281.
Croson, Rachel T.A., 1996, "Partners and Strangers Revisited." *Economic Letters*, *53*, 25-32.
----- and Buchan, Nancy, 1999, "Gender and Culture: International Experimental Evidence from Trust Games," *American Economic Review*, *89*(2), 386-391.
Dickinson, David L., 2001, "The Carrot Vs. the Stick in Work Team Motivation," *Experimental Economics*, *4*(1), 107-124.
Engelmann, Dirk, and Strobel, Martin, 2004, "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," *American Economic Review*, *94*(4), 857-869.
Falk, Armin and Fischbacher, Urs, 2000, "A Theory of Reciprocity," University of Zurich, (http://www.iew.unizh.ch/wp/iewwp006.pdf)
Falk, Armin, Fehr, Ernst and Fischbacher, Urs, 2003, "On the Nature of Fair Behavior," *Economic Inquiry*, *41*(1), 20-26.
Falk, Armin, Fehr, Ernst and Fischbacher, Urs, "Testing Theories of Fairness – Intentions Matter," 2000, Working Paper No. 63, Institute of Empirical Research in Economics, University of Zurich.
Fehr, Ernst and Fischbacher, Urs, 2004, "Third Party Punishment and Social Norms," *Evolution and Human Behavior*, 25, 63-87.
Fehr, Ernst and Schmidt, Klaus M., 1999, "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, *114*(3), 817-868.
Forsythe, Robert, Horowitz, Joel L., Savin, N. E. and Sefton, Martin, 1994, "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, *6*, 1994, 347-369.
Frank, Robert, 1988, *Passions within Reason: The Strategic Role of the Emotions*. London: Norton.
Gächter, Simon and Riedl, Arno, 2005, "Moral Property Rights in Bargaining," *Management Science*, 51(2), 249-263.
Güth, Werner and Ockenfels, Axel, 2000, "Evolutionary Norm Enforcement," *Journal of Institutional and Theoretical Economics*, 156(2), 335-347.
Güth, Werner, Schmittberger, Rolf and Schwarze, Bernd, 1982, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, *3*, 367-388.
Hoffman, Elizabeth, McCabe, Kevin, Shachat, Keith, and Smith, Vernon, 1994, "Preferences, Property Rights and Anonymity in Bargaining Games," *Games and Economic Behavior*, *7*, 346-380.
Kachelmeier, Steven J., Stephen T. Limberg, and Michael S. Schadewald. 1991. "A Laboratory Market Examination of the Consumer Price Response to Information About Producers' Costs and Profits." *The Accounting Review* 66:4, pp. 694-717.

Kahneman, Daniel, Knetsch, Jack L. and Thaler, Richard, 1986a, "Fairness and the Assumptions of Economics," *Journal of Business*, *59*:4, S285-S300.
Kahneman, Daniel, Knetsch, Jack L. and Thaler, Richard, 1986b, "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review*, *76*, 728-741.
Konow, James, 2005, "Blind Spots: The Effects of Information and Stakes on Fairness Biases," *Social Justice Research*, *18*(4), 349-390.
Konow, James, 2000, "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions," *The American Economic Review*, *90*(4), 1072-1092.
Konow, James, 2003, "Which is the Fairest One of All?: A Positive Analysis of Justice Theories," *Journal of Economic Literature*, *41*(4), 1186-1237.
Levine, David K., 1998, "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, *1*, 593-622.
McCabe, Kevin A., Rigdon, Mary L. and Smith, Vernon L., 2003, "Positive Reciprocity and Intentions in Trust Games," *Journal of Economic Behavior and Organization*, *52*, 267-275.
Offerman, Theo, 2002, "Hurting Hurts More Than Helping Helps," *European Economic Review*, *46*, 1423-1437.
Rabin, Matthew, 1993, "Incorporating Fairness Into Game Theory and Economics," *American Economic Review*, *83*(5), 1281-1302.
Rutström, E. Elisabet, and Williams, Melonie B., 2000, "Entitlements and Fairness: An Experimental Study of Distributive Preferences," *Journal of Economic Behavior and Organization*, *43*, 75-89.
Smith, Adam, 1759 [1809], *The Theory of Moral Sentiments*, Glasgow: R. Chapman.