

# Fairness and the Road Not Taken: An Experimental Test of Non-Reciprocal Set-Dependence in Distributive Preferences

Martin Eiliv Sandbu

Columbia University

sandbu@post.harvard.edu

July 2004

## Abstract

Experimental investigations of preferences for fairness have revealed systematic set-dependence in people's allocative choices: Choices over identical options can be reversed depending on the larger game within which the choice is embedded. Some of these reversals have been shown to reflect *reciprocity motives*, the desire to help those who help you and hurt those who hurt you. This paper highlights the paucity of investigations of *non-reciprocal set-dependence*. It documents choice reversals that cannot be explained by the desire to reciprocate intentions, and proposes a theory that allows for both reciprocal and non-reciprocal set-dependence. In this theory, the weight an agent puts on other people's payoffs depends on how much those people are thought to deserve, which in turn may be affected both by reciprocity and directly by the set of available allocative outcomes. The theory is tested against experimental data to show that non-reciprocal set-dependence is as quantitatively important as reciprocity.

# 1 Introduction

One of the great contributions to economic science in recent years has been the investigation of social motivations in economic choices. Experimental researchers have accumulated convincing evidence that people's decisions depend not only on their own payoffs but on the payoffs of others (see Camerer (2003) for a comprehensive review). The laboratory research has started to be complemented by theoretical investigations of other-regarding preferences: There is now a sizeable literature on "social preferences" or utility functions defined over distributions of payoffs (see in particular Rabin 1993, Bolton and Ockenfels 2000, Andreoni and Miller 2002, Charness and Rabin 2002). Andreoni and Miller (2002, hereafter AM) in particular contribute to reconciling the evidence with theory by showing that altruistic behaviour in a dictator game experiment conforms to the consistency conditions of the generalised axiom of revealed preference (GARP).

One lesson about other-regarding preferences that can be drawn from the experimental literature is that fairness is not simply a function of the distributive outcome of a game. Put differently, typical choices among payoff distributions cannot be rationalised by an exclusively "outcome-based" utility function, defined only over payoff distributions. This is because allocative choices exhibit set-dependence – they are found to vary systematically with the possible outcomes off the path that is actually played, even in the final-stage subgames, where the set of originally available outcomes should be irrelevant in outcome-based theories. (Prasnikar and Roth 1992, Blount 1995, Güth, Huck and Müller 2001, Andreoni, Brown and Vesterlund 2002, Falk, Fehr and Fischbacher 1999)

The principal explanation that has been proposed for set-dependent behaviour has been reciprocity theory (Rabin 1993, Dufwenberg and Kirchsteiger 1998, Falk and Fischbacher 2000). Reciprocity models assume that agents are motivated to help those who have helped them and vice-versa. This assumption rationalises set-dependence insofar as the set of available outcomes affects what intentions can be inferred from a given action by another player. Still, reciprocity is only one of many possible factors that could lead to choice reversals. In section 2 I offer example of how the set of available outcomes affects people's behaviour even in the absence of any actions

by other players. This suggests that people's perceptions of fairness are shaped by the available set through mechanisms other than reciprocity motives.<sup>1</sup>

The obvious plausibility of reciprocity theory and its ability to explain many behavioral regularities is paralleled by a neglect of other causes of set-dependent choice reversals.<sup>2</sup> As a result, there are few experiments in which reciprocal behaviour is not observationally equivalent with non-reciprocal set-dependence.<sup>3</sup> This paper explores whether reciprocity theory is sufficient to explain choice reversals and set-dependent behaviour. It does so by investigating experimental games in which agents could exhibit set-dependence without exhibiting reciprocity, and measures the importance of both reciprocity and non-reciprocal set-dependence. In section 3, I present a simple utility function that is calibrated to experimental data. In section 4 I demonstrate the presence of set-dependence in unilateral decision problems where there can be no reciprocity motives. In section 5 I use data from 28 experimental games, due to Charness and Rabin (2002, hereafter CR), to compare the quantitative importance of non-reciprocal set-dependence and reciprocity motives. When I generalise CR's model in order to allow for non-reciprocal set-dependence, I show that the latter is at least as important as reciprocity motives in explaining behaviour. Section 6 verifies that the results are not simply due to a phenomenon known in the marketing literature as "context-dependence," and section 7 concludes.

## 2 Intuitions: The relevance of irrelevant alternatives

Güth, Huck and Müller (2001) present a particularly clear demonstration of set-dependent choice reversals over identical pairs of distributive outcome. Consider the three games displayed in figure 1. In each of these games, Ann chooses between a (17,3)-offer<sup>4</sup> in her favour and another offer

---

<sup>1</sup>Moreover, even when it is previous actions that affect current preferences they need not do so through reciprocity. Promise-keeping (?) and truth-telling (Brandts and Charness 1999) are two examples of non-reciprocal motives that have received attention by experimental investigators.

<sup>2</sup>Occasionally the set-dependence is treated as *equivalent* to reciprocity, as when behaviour that is sensitive to variations in the available set of outcomes is interpreted as a proof that "intentions matter" (Falk, Fehr and Fischbacher 1999).

<sup>3</sup>One exception is the set of experiments reported by Charness and Rabin (2002). Even though this was not part of their study, their experimental data set is sufficiently rich to test for the presence of both reciprocal and non-reciprocal set-dependence. I perform this analysis on their data in section 5.

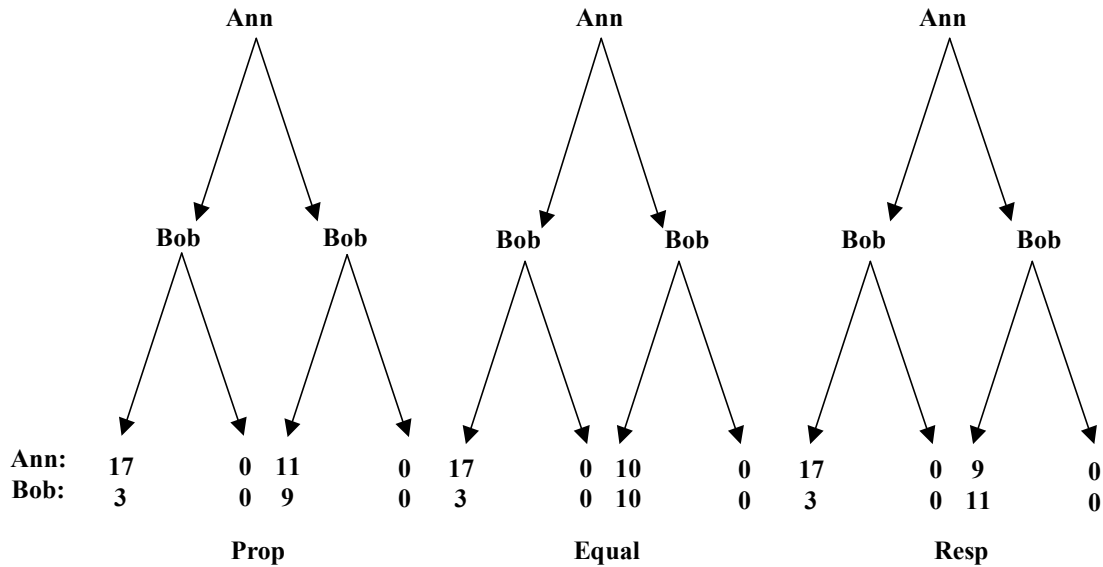


Figure 1: Three mini-ultimatum games with equal or almost equal offers, due to Güth, Huck and Müller (2001).

which varies across the games, then Bob decides whether to accept or reject. If he rejects the offer both players get zero. In game “Equal,” the second available offer is an equal split (10,10). In the other two games the second allocation is slightly unequal, either favouring the proposer Ann (game “Prop”) or the responder Bob (game “Resp”). The standard selfish model predicts that all offers should be the very unequal (17,3) and that it should always be accepted. In fact, more than half of the proposers make the fair offer, and almost half of the responders reject the unfair one. But the most surprising result in this experiment was the variation in behaviour *across* the three games. Proposers make the fairer offer 70% of the time when it is exactly egalitarian, but choose it considerably less often when it is slightly unequal, especially when the slight inequality favours the responder (55% in Prop and only 33% in Resp). The responders’ frequency of rejecting the unfair (17,3)-offer is especially low (25%) when all agreements favour the proposer, whereas it is over one-half when the fair offer is equal or favours the responder (60% in Equal and 50% in Resp).

Other studies document the same sensitivity to strategically irrelevant alternatives. Falk, Fehr

---

<sup>4</sup>The payoff unit was Deutsche Mark.

and Fischbacher (1999), in another mini-ultimatum game experiment, report similar findings with much larger inequalities in the fairest offers. Andreoni, Brown and Vesterlund (2002) study the phenomenon in a more complex experiment. They conclude that “if models of fairness are to predict the observed difference across... games, then they must allow the evaluation of actions to depend on the actions not chosen... Not only is the actual allocation producing fairness, but the road to that allocation and the roads not taken along the way are also inputs into the production of fairness.”

The importance of the “road not taken” has been formally modelled by reciprocity theories (Rabin 1993, Dufwenberg and Kirchsteiger 1998, Falk and Fischbacher 2000). In these models, preferences over distributive outcomes can be reversed because “intentions matter.” People are motivated to reward those who behave kindly or fairly towards them and to punish those who behave unkindly or unfairly. The theory has intuitive appeal, and serves to explain such well-known phenomena as rejections in ultimatum games (see Roth 1995) and the willingness to punish people who do not contribute to public goods (see Fehr and Gächter 2002). It is important, however, to notice the conceptual difference between set-dependence and reciprocity. The almost exclusive focus on reciprocity theory as the explanation of set-dependence risks giving the impression that reciprocity theory is by definition the only alternative to purely outcome-based theories. In fact reciprocity is only one of many things that could induce preference changes over an indential pair of distributive outcomes. There is a difference between saying that *the set of alternatives* matters and that *intentions* matter. The former does not entail the latter, although observations of the former are sometimes interpreted as evidence for the latter. At least one study, on the other hand, claims to show that it is not intentions that matter. Bolton, Brandts and Ockenfels (1998) reports an experiment in which the second mover was given the same range of choices in three different conditions. The choice was put to the player conditional on an “unkind” move by the first mover, a “kind” move by the first mover, or no move by the first first mover. They found little of the difference between conditions that reciprocity predicts there to be.

To gain intuition for why the set of alternatives can matter even in the absence of intentions to

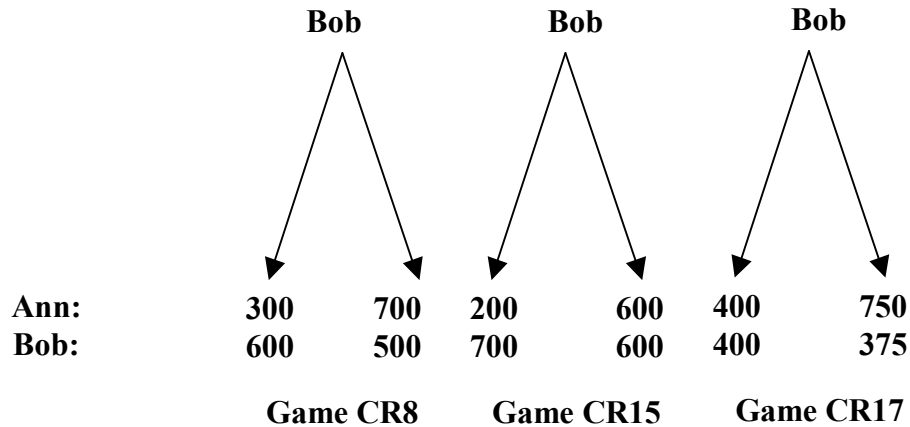


Figure 2: Three individual decision problems illustrating set-dependence. Due to Charness and Rabin (2002).

reciprocate, it is useful to consider individual decision problems, where by definition there are no actions by the other player to reciprocate. The three degenerate games in figure 2 are games 8, 15 and 17 from the series of experiments analysed by Charness and Rabin (2002), which I discuss in detail in section 5: In the first two games, Bob has the opportunity to sacrifice 100 points<sup>5</sup> in order to improve Ann’s payoff by 400 points. CR found that in the first situation, 33% of the subjects playing the role of Bob sacrificed 100 points by choosing the right-hand branch which increases Ann’s payoff. In the second situation, however, a full 73% of the subjects chose to sacrifice. This dramatic difference cannot be due to reciprocity motives, since there is no action performed by Ann for Bob to reciprocate. Can it be accounted for by purely outcome-based models? One rationalisation of the difference in the frequency of sacrifice could be that people are behindness-averse, that is, they are prepared to sacrifice to help others when they are ahead, but do not want to end up behind. The problem with this explanation is that CR find no evidence of behindness-aversion in their sample of games. The game CR17 provides an illustration. The right-hand branch in this situation leads to an even more unequal allocation in Bob’s disfavour than in CR8. Yet 50% of the subjects choose the (750, 375)-allocation in CR17, considerably more than choose the (700, 500)-allocation in CR8. When CR fit utility functions for the entire sample of games, they find no

<sup>5</sup>100 experimental points were worth 100 pesetas (about 70 cents at the contemporaneous exchange rate) in game CR8 and \$1 in CR15 and CR17.

overall aversion to being behind.

What is behind these set-dependent choice reversals? In CR8 there is no available egalitarian allocation, in contrast to the other two situations where (600, 600) and (400, 400), respectively, are possible outcomes. This suggests a pattern of non-reciprocal set-dependence in which the presence of more or less fair allocations influences the willingness to sacrifice for the benefit of the other player.<sup>6</sup> When there is no opportunity for equality, as in CR8, distributive fairness may simply seem less important — since no choices will actually achieve it. Ann cannot fairly claim an equal share since an equal share is not available. In contrast, an unequal allocation is perceived as more unfair in a situation where an egalitarian allocation is available, such as in CR15 and CR17. The availability of equality makes Ann’s claim to payoffs more salient. In the rest of the paper, I investigate this intuition more systematically. The remainder of this section presents a formal model which can incorporate both reciprocity and non-reciprocal set-dependence. In the subsequent sections I calibrate it to experimental data in order to establish the presence of non-reciprocal set-dependence and to measure the relative importance of the two patterns.

## 3 Formalisations

### 3.1 Modelling distributive preferences

We know from the experimental research that people care about fairness (or behave as if they do): Their choices depend not only on the consequences for their own payoffs, but for distributions of payoffs among vary individuals. Purely outcome-based models, however, are not sufficient to capture that behaviour. This is not surprising inasmuch as there is more to fairness than outcome fairness. Whether a distributive outcome is fair or not depends on whether it gives everyone what they deserve, so an equal distribution, say, is only the fairest if every person is equally deserving. The function from outcomes to fairness may therefore depend on *how* the outcome comes about.

---

<sup>6</sup>The possibility that the value of an option depends on the menu of options from which it is selected is not limited to the case of distributive preferences. A profound treatment of this phenomenon can be found in Jon Elster’s book *Sour Grapes* (Elster 1983). Some formal characteristics of menu-dependence are analysed by Sen (1997).

Consider three situations: (a) two friends find a sum of money in the street; (b) two poker players freely decide to enter a high-stake poker game; (c) money has to be divided among two individuals, one of whom has hurt the other before. We can expect that if we ask people to evaluate the relative fairness of an identical set of distributions most would give different answers for each of the three situations.<sup>7</sup>

These intuitions can be captured in the following simple 2-person model of distributive preferences:<sup>8</sup>

$$U_{\mathbf{r}}(\mathbf{x}) = \text{sign}(\rho) \left[ (1 - \alpha(\mathbf{r})) x_{self}^{\rho} + \alpha(\mathbf{r}) x_{other}^{\rho} \right], \quad (1)$$

The utility function  $U_{\mathbf{r}}$  ranks vectors  $\mathbf{x} \in \mathbb{R}_+^2$  which denote allocations of payoffs to each individual:  $\mathbf{x} \equiv (x_{self}, x_{other})$  where  $x_{self}$  is the payoff to the decision-maker. It captures the sensitivity of altruism to non-outcome factors by letting the weights depend on a *fair reference payoff vector*, denoted  $\mathbf{r} \equiv (r_{self}, r_{other}) \in \mathbb{R}_+^2$ .

It will be convenient to write the utility function in a weighted-sum specification as well as the weighted-average specification. As long as the individual puts positive weight on her own payoff ( $\alpha < 1$ ), the same ordinal preferences as in utility function (1) can be represented by:

$$U_{\mathbf{r}}(\mathbf{x}) = \text{sign}(\rho) \left[ x_{self}^{\rho} + A(\mathbf{r}) x_{other}^{\rho} \right] \quad (2)$$

with  $A = \alpha / (1 - \alpha)$ . Note that  $A$  is the marginal rate of substitution between own and other's payoffs at egalitarian allocations ( $x_{other}/x_{self} = 1$ ). I have shown elsewhere (Sandbu 2003) that this utility function can be derived from a simple axiomatic basis which also entails that  $A(\mathbf{r})$  is

---

<sup>7</sup>Kahneman, Knetsch and Thaler (1986*b,a*) give more examples of how people judge the fairness of payoff allocations differently depending on the context in which the outcome arises.

<sup>8</sup>The family of CES functions defined by different values of  $\rho$  should be understood to include its limit case as  $\rho \rightarrow 0$ , the Cobb-Douglas functional form:

$$U_{\mathbf{r}}(\mathbf{x}) = (1 - \alpha(\mathbf{r})) \ln x_{self} + \alpha(\mathbf{r}) \ln x_{other}$$

homogeneous of degree zero in  $\mathbf{r}$ .<sup>9</sup> The simplest form for the  $A(\mathbf{r})$ -function is:

$$A(\mathbf{r}) = a + cr \quad (3)$$

where  $r$  denotes the ratio of reference payoffs,

$$r \equiv \frac{r_{other}}{r_{self}},$$

and  $a$  and  $c$  are constant scalars. (Throughout the paper, greek letters are used as coefficients in the weighted average specification and latin letters in the weighted sum specification).

This utility function defines a trade-off between own payoff and payoff to the other person that is parametrised by  $\rho$ ,  $a$ , and  $c$ . The elasticity of substitution  $1/(1 - \rho)$  measures the sensitivity to inequality (the lower is the curvature parameter  $\rho$ , the more elastic is the MRS with respect to changes in the payoff ratio). In the limit when  $\rho \rightarrow -\infty$ , the utility function approaches the Leontief form. The weight on the other persons payoff may depend on the reference payoff ratio  $r$ . The sensitivity of the MRS to the reference payoff ratio at equality is given by  $c$ , which is therefore a measure of reference-dependence. When  $c = 0$ , the reference payoffs have no influence on preferences, and the weight on the other person's payoff is given by  $a$ , which can therefore be seen as a measure of pure (reference-independent) altruism.

To generate testable predictions, the theory naturally needs to specify possible determinants of the reference point. Different of fairness motivation can then be differentiated by the way they model those determinants. In a merely outcome-based theory,  $\mathbf{r}$  would just be a constant. In a reciprocity theory, the reference point would be a function of the strategies played by the players and their available strategy sets:  $\mathbf{r} = \mathbf{r}(s_{self}, s_{other}, \mathbf{S}_{self}, \mathbf{S}_{other})$ , where  $\mathbf{S}_i$  denotes the set of available strategies for player  $i$ . In the case of non-reciprocal set-dependence, the reference-point would be a direct function of the set of available payoff allocations:  $\mathbf{r} = \mathbf{r}(\mathbf{X})$ , where  $\mathbf{X}$  denotes the

---

<sup>9</sup>The axiomatic treatment imposes conditions of separability, non-discrimination, and homotheticity on the utility function  $U(\bullet)$  and on the reference-dependence function  $A(\bullet)$ . Detailed definitions and proofs can be found in Sandbu (2003).

set of all possible payoff allocations available in the game under considerations.<sup>10</sup> In what follows I elaborate and test a theory of (non-reciprocal) set-dependent reference points and compare it with reciprocity motives in experimental data.

### 3.2 Modelling non-reciprocal set-dependence

A simple way of conceptualising non-reciprocal set-dependence is to assume that the decision-maker's reference allocation is what she perceives as the fairest point in the set of all available payoff allocations. There are three criteria it is natural to try to reconcile in the choice of such a reference allocation:

1. *Egalitarianism*: There is a presumption that fairness involves treating everyone equally. The equal split has a particularly salient role in distributive questions, and the reference allocation should be one which avoids excessive inequality.
2. *Efficiency*: How much individuals deserve should depend on the opportunity costs of allocating payoffs to them. If someone can be benefited only at the cost of large sacrifices from someone else, they may have a weaker claim than if their position can be improved at negligible cost.
3. *Legitimate self-interest*: Fairness does not exclude the pursuit of self-interest altogether. In particular, fairness does not require helping someone very disadvantaged if that can only be done by taking that person's position. Fairness may demand that the rich help the poor and diminish the inequality between them; but not that the rich take the place of the poor. For an example closer to laboratory choices, consider a decision-maker who has to choose between \$8 for herself and \$2 for the other person, versus only \$2.10 for herself and \$7.90 for the other person. Fairness does not require her to make this personal sacrifice for such a small reduction in inequality (and no improvement in efficiency).<sup>11</sup>

---

<sup>10</sup>These three suggestions do not, of course, exhaust the possible determinants of the reference point. A model aiming for full descriptive realism might include such factors as status (Cox and Friedman 2002 discussed in ), gender (Andreoni and Petrie 2004 investigated by ), or property rights or entitlements (Gächter and Riedl 2002).

<sup>11</sup>A similar criterion is incorporated in Falk and Fischbacher's (2000) theory of reciprocity. In their model, however,

The trade-off between equality and efficiency is well-known. A compromise between them will allow some inequality for the benefit of increasing absolute payoffs, such as in the “lexicographic maximin” or “leximin” allocation. This is the allocation that gives as much as possible to the person who gets the least, then conditional on that, as much as possible to the second worst-off person, and so on. The leximin allocation is Pareto efficient, and in the two-person case it is the most egalitarian of all the efficient allocations. The leximin allocation is therefore a good candidate for determining reference payoffs. Indeed, John Rawls (1971) argues that the leximin allocation is what rational parties to a social contract would agree on if they did not know their backgrounds and positions in society.<sup>12</sup> Behind Rawls’s “veil of ignorance” a decision-maker cannot pursue her self-interest except by pursuing the interests of whoever she may turn out to be. In the kinds of situations we are investigating, however, agents typically do know their position, and the leximin allocation may fall foul of the legitimate self-interest criterion. A fair reference allocation which also permits self-interest in the two-person case is the leximin allocation *among those allocations that do not put the decision-maker in a disadvantaged position*. This rule for fair reference allocations respects equity and efficiency concerns, but does not admit reference allocations that would put the decision-maker behind (unless there are *only* such allocations).

Formally, denote by  $\mathbf{X} \equiv \{\mathbf{x} \in \mathbb{R}_+^2 \mid \mathbf{x} \text{ is available}\}$  the set of available payoff distributions (a distribution is available if it is achievable through some combination of actions by the agents involved). Denote by  $\mathbf{X}^{advantaged}$  the subset of  $\mathbf{X}$  whose elements give the decision-maker at least as much as the other person, that is,  $\mathbf{X}^{advantaged} \equiv \{\mathbf{x} \in \mathbf{X} \mid x_{other} \leq x_{self}\}$ . The fair reference allocation  $\mathbf{r} \equiv (r_{self}, r_{other})$  is a function of  $\mathbf{X}$  characterised by:

- If  $\mathbf{X}^{advantaged}$  is non-empty,  $\mathbf{r}(\mathbf{X})$  is the payoff vector  $\mathbf{r}$  in  $\mathbf{X}^{advantaged}$  that satisfies:

1.  $r_{other} \geq x'_{other}$ , for all  $\mathbf{x}' \in \mathbf{X}^{advantaged}$ , and

---

legitimate self-interest enters the model as a determinant of the kindness or fairness of an *action* or *intention* to act. In the model presented here, on the other hand, the criterion is at work even when there are no actions or intentions to reciprocate.

<sup>12</sup>Rawls gives absolute priority to principles of equal liberties and opportunities before other distributive questions are dealt with. Moreover, his leximin requirement (the "difference principle") applies not to incomes and wealth directly, but to an "index of primary goods." Within these provisos, however, leximin is an important principle in Rawls's theory and an important corrective to utilitarianism's light-handedness with distributive fairness.

2.  $r_{self} \geq x''_{self}$ , for all  $\mathbf{x}'' \in \mathbf{X}^{advantaged}$  satisfying (1).

- If  $\mathbf{X}^{advantaged}$  is empty,  $\mathbf{r}(\mathbf{X})$  is the payoff vector  $\mathbf{r}$  in  $\mathbf{X}$  that satisfies:

3.  $r_{self} \geq x'_{self}$  for all  $\mathbf{x}' \in \mathbf{X}$ , and

4.  $r_{other} \geq x''_{other}$  for all  $\mathbf{x}'' \in \mathbf{X}$  satisfying (3).

With this definition we can proceed to measuring the importance of non-reciprocal set-dependence in laboratory behaviour.

## 4 Measuring set-dependence

### 4.1 Modifying Andreoni and Miller (2002)

The obvious way to isolate the set-dependence effect from any reciprocity effects is to look for set-dependence in individual decision problems, where reciprocity can play no role. This section reports a dictator game experiment constructed to investigate the set-dependence hypothesis. The design was based on an experiment by Andreoni and Miller (2002, hereafter AM). AM offered subjects a unilateral choice of how to divide a series of budgets between themselves and another participant, where the exchange rate between payoff to self and payoff to other could take on different values. They found that subject behaviour was well described by a CES function defined over own and other's payoff:

$$U(\mathbf{x}) = \left[ (1 - \alpha)x_{self}^\rho + \alpha x_{other}^\rho \right]^{\frac{1}{\rho}} \quad (4)$$

which describes the same behaviour as the general utility function (2,3) with  $c = 0$ . The authors identified three ideal-type preferences: Preferences that are selfish ( $\alpha = 0$ ), egalitarian ( $\rho \rightarrow -\infty$ ) or perfect substitutes-type ( $\rho = 1$ ). The behaviour of about half of the subjects conformed exactly to one of the three ideal types. The remaining half were sorted as “weak” versions of the three strong types depending on how close their choices were to each of those strong types (closeness

was measured by the Euclidian distance in payoff space), and a utility function was estimated for each weak type from pooled choice data for each group.

My experiment builds on AM's design but modifies it so as to measure the influence of the available set on the distributive choices of the subjects. By changing the feasible set I vary the reference point and examine how subject behaviour changes as result. The experimental design is explained in detail in the next subsection. I follow AM's approach of estimating types of utility functions based on subjects' choices in dictatorial division tasks, but I widen the range of decision problems faced by the subjects so as to manipulate the shape of the budget set. Like AM, I treat each of the six types as internally homogeneous. The strong types are perfectly characterised by the utility functions described in the previous paragraph, while I estimate the parameters for the three "weak" types. The three ideal types by definition do not exhibit set-dependence. The analysis therefore focuses on whether the preferences of the weak types (roughly two-thirds of the subject) are set-dependent.

My benchmark group consists of the weakly selfish individuals, which is the most numerous subgroup (45 out of a total of 63 weak type subjects), but I also make allowance for the other weak two types in the estimation from pooled data by including dummy variables in the altruism weight and the curvature parameter. This is done through the following estimating equation, derived from the first-order condition of the utility function:<sup>13</sup>

$$\frac{x_{other,ij}}{x_{self,ij} + x_{other,ij}} = \frac{A_{ij}^{\frac{1}{1-\rho_i}}}{p_j^{\frac{1}{1-\rho_i}} + A_{ij}^{\frac{1}{1-\rho_i}}} + \varepsilon_{ij}, \quad (5)$$

with

$$A_{ij} = a + cr_j + b_E WEAKEGAL_i + b_{PS} WEA KPSUB_i \quad (6)$$

$$\rho_i = g + g_E WEAKEGAL_i + g_{PS} WEA KPSUB_i \quad (7)$$

---

<sup>13</sup>The derivation of the estimating equation is shown in the appendix.

where  $i$  indexes the subject and  $j$  indexes the decision problem. The dependent variable is the share of the pie given by subject  $i$  to the recipient in decision problem  $j$ .  $A_{ij}$  is the econometric operationalisation of  $\alpha(\mathbf{r}) / (1 - \alpha(\mathbf{r}))$  in the theoretical models (cf. equations 1, 2 and 3); it is the marginal rate of substitution at equality of subject  $i$  given an available set of allocations with fair reference point  $j$ .  $\rho_i$  is the curvature of the utility function of subject  $i$ , and like the altruism weight is homogeneous within subject types.  $p_j$  is the opportunity cost of giving in game  $j$ . The indicator variable WEAKEGAL is set to one if and only if the subject is classified as weakly egalitarian and zero otherwise, WEAKPSUB is set to one if and only if the subject is classified as weakly perfect-substitutes and zero otherwise, and  $\varepsilon_{ij}$  is a normally distributed observation disturbance which we allow to be correlated within but not across subjects. The variable of interest for the set-dependence hypothesis is the coefficient on  $r$ , which is the reference payoff ratio described in subsection 3. Set-dependence predicts that the coefficient  $c$  should be statistically significant and positive (a higher reference payoff for the other player should lead the decision-maker to put a more positive weight on that player's actual payoff).<sup>14</sup> The estimation uses maximum likelihood, where the likelihood function is a tobit model which treats corner choices as censored observations. This prevents the coefficient  $c$  from picking up any effects of changes in the available set on choice behaviour that simply reflect that the most preferred options become unavailable when the set is restricted. The standard errors are adjusted for clustering on subjects.

## 4.2 Experimental design

The experiment tests for set-dependence by adding one important feature to the AM design: The budget sets are truncated. The truncations were designed so as to investigate whether the availability of more or less fair allocations affects preferences, as set-dependence predicts. There were a total of 36 decisions to be made. To vary the opportunity cost of giving, there were four different

---

<sup>14</sup>Note that  $r$  is allowed to influence the altruism weight  $A$  but not the curvature parameter  $\rho$ . This respects the theoretical foundations of the model, which include an axiom of independence of irrelevant reference payoffs (individual separability in reference payoffs) that rules out a dependence of  $\rho$  on  $r$  in the multi-player case. The intuition is that the preference over two allocations that differ only in the payoffs to two individuals should not be affected by changes in the reference payoff of a third individual who gets the same in both allocations. See Sandbu (2003) for details.

exchange rates between the dividers' and the recipients' payoffs, making the smallest possible pie \$12 and the largest possible \$42 (only achievable if the divider gave everything to the recipient). The nine types of truncation are schematically illustrated in figure 3. For each exchange rate, three

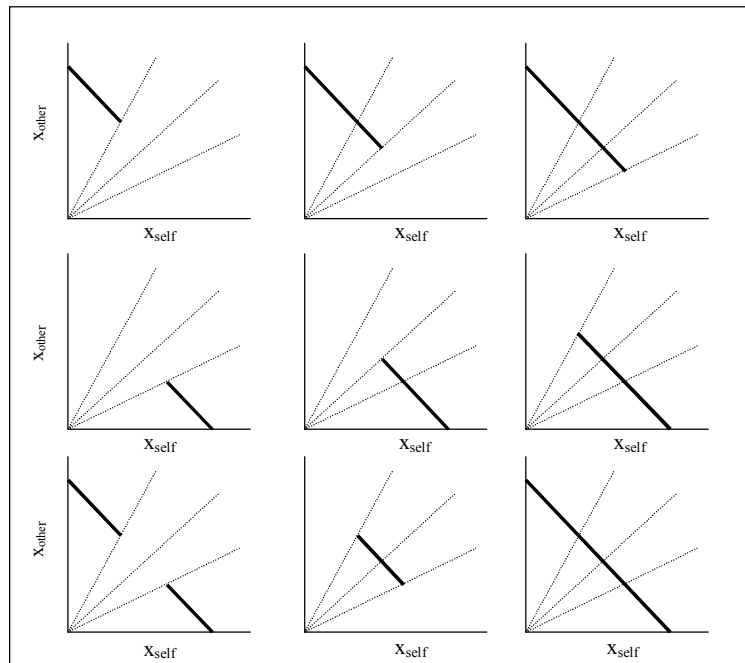


Figure 3: Truncation of budget sets in the dictator game experiment

questions had budgets truncated from above, three questions had budget sets truncated from below, and three questions had symmetric budgets. The asymmetrically truncated budgets ensured that either the divider (when the truncation was from below, cf. row 1) or the recipient (when the truncation was from above, cf. row 2) could get at most  $1/3$ ,  $\frac{1}{2}$  or  $2/3$  of the aggregate dollar earnings in the division, respectively. The symmetric budgets (cf. row 3) included one unrestricted budget (the divider could give any amount between zero and everything), one which required that both participants should get at least  $1/3$  of the pie, thus limiting the maximum amount of inequality, and one which required that one of the two participants (the divider's choice) should get at least  $2/3$ , thus making an equal division unavailable.

The questions given to the subjects specified which kinds of divisions were allowed in each case. Two typical questions might read:

“Divide 120 tokens. Each token is worth 10 cents to you, 20 cents to the other person.  
You must pass at least 20 tokens and at most 60 tokens.”

“Divide 120 tokens. Each token is worth 10 cents to you, 10 cents to the other person.  
You must pass less than 40 tokens or more than 80 tokens.”

For each question, the subjects were asked to enter the number of tokens they passed<sup>15</sup> to the other person, after which the computer would show the monetary distribution implied by the decision, then give the subjects a chance to change their minds. Before finalising each decision, the subjects had an unlimited opportunity to go back and change the number of tokens they gave. This procedure minimised the risk that confusion and mathematical complexity prevented the subjects from acting on their preferences. Once the subjects had confirmed their decision, they would move on to the next choice situation. Each subject was presented with the same 36 decision problems, but in different (random) orders. This was common knowledge. The subjects were told that one of their decisions would be chosen at random at the end of the study, and it would not be the same choice situation for each subject.<sup>16</sup> They made money from what they had decided to keep in the chosen decision problem, and from being a recipient from somebody else’s division. The recipients were also paid a show-up fee of \$10.<sup>17</sup>

The experiment was programmed and conducted using the zTree software for economic experiments (Fischbacher 1999). It was carried out in the course of three sessions, which included a

---

<sup>15</sup>We followed AM in using this terminology to avoid connotations of giving and charity. It also makes our results more directly comparable with theirs.

<sup>16</sup>Earlier readers of this paper have expressed worries about this way of rewarding the subject. Since the recipient cannot in any way affect the payoffs of the divider he or she is matched with, there is of course no direct reciprocity here. But the fact that everyone is both a divider and a recipient (with different participants) it may create *indirect* reciprocity motives. That is, a subject may expect to be treated generously/stingily by his or her divider when in the recipient role, and will “indirectly reciprocate” in the divider role. But it should be noted that even if there are indirect reciprocity motives, they could at most affect the constant  $a$  in the altruism weight, and not its sensitivity  $c$  to the reference point. How the subjects expect to be treated by a randomly selected other subject in a randomly selected other decision problem does not vary across the decisions they themselves have to make. Therefore, the influence of indirect reciprocity should be the same in all the decisions (and should show up in the estimate of  $a$ ). A significant estimate of  $c$ , therefore, cannot be explained by indirect reciprocity motives.

<sup>17</sup>The detailed instructions are available from the author as an appendix to this paper. The reported 36 dictator games were the first part of a two-part experiment. The second part was an anonymous two-player sequential-move game, which is immaterial for the present paper. The only thing that the subjects knew during the first part of the experiment was that there would be a second part where they would anonymously interact with a *new* randomly matched participant and that they would have a possibility of earning more money.

total of 96 subjects. I calibrate the utility functions on the behaviour of those 65% of the subjects who are classified as “weak” types. I therefore restrict the sample to 63 subjects. With 36 decisions made by each person, this gives me 2268 observations. Since even the “weak” types sometimes change their choice not because of set-dependence but because their (unchanged) favourite option becomes unavailable, I use a tobit estimation procedure which takes into account the truncation of the budget sets and treats corner choices as censored choices. This avoids interpreting pure truncation effects as set-dependence.

### 4.3 Results

The subjects in my experiment are distributed in roughly the same way as AM’s subjects.<sup>18</sup> 33 out of 96 subjects are “strong” types whose behaviour is exactly captured by the relevant utility function. They clearly do not exhibit any set-dependence. If these subjects alter their behaviour when the available set changes, that is only because their previous choice becomes unavailable and they are forced to choose their best possible corner solution. I therefore look for set-dependence in the behaviour of the two-thirds of subjects that at least once chose an interior solution (and that did not exactly fit perfectly egalitarian or perfect substitutes preferences). An example will illustrate how preferences can be properly characterised as set-dependent. Consider the three budget sets depicted in the bottom row of figure 3, which we may call an “outer,” “inner,” or “spanning” budget, respectively. A particularly strong example of set-dependence can be found in one subject’s choices when the price of giving was 1/3. In these situations, the subject was asked to divide 140 tokens with a recipient, with each token being worth 10 cents to the divider and 30 cents to the recipients. In the “outer” budget, the subject had to pass no more than 20 or no fewer than 56 tokens, and chose to pass 20 (a \$12-\$6 division in the divider’s favour). In the “inner” budget, however, where the number of tokens passed had to be between 20 and 56, the subject chose 56

---

<sup>18</sup>65% of all the subjects (63 out of a total of 96) were weak types, which is somewhat more than what AM found. This is not surprising, since my subjects faced many more choice situations and thus had more opportunity to make “imperfect” choices. Of the 63 weak subjects, 45 or 71% were weakly selfish, 17 or 27% were weakly egalitarian, and only one subject was a weak perfect substitutes-type. Among the other 33 subjects, one was a strong perfect substitutes-type, and the remaining 32 were all strongly selfish. Overall, this means my pool had somewhat more selfish types and fewer perfect substitute-types than AM.

(a \$16.80-\$8.40 division in the recipient’s favour). Finally, in the “spanning” budget, where any division was allowed, the subject chose to pass 60 tokens (a \$18-\$8 division in the recipient’s favour). These choices clearly violate the standard consistency axioms *if* preferences are only defined over distributive outcomes<sup>19</sup>. In the set-dependence theory, however, this “reversal” is easily explained: In the latter two budgets the most fair allocation gives a larger payoff to the recipient, whose fair claim is therefore higher. This in turn tilts the divider’s preferences to the recipient’s advantage.

Such examples very clearly bring out the kind of behaviour that the set-dependence theory attempts to capture. Single examples do not, of course, provide a comprehensive test of its hypotheses. I therefore proceed to the econometric analysis of the entire sample in which I calibrate the set-dependent utility function (2, 3) to the behaviour of all the weak type-subjects in order to quantify the general importance of set-dependence. Table 1 gives the parameter estimates for three different models. The first is a set-independent model where the coefficient on  $r$  is constrained to be 0 (model DG1), so the only estimated parameters are  $\hat{a}$ ,  $\hat{\rho}$ , and the coefficients on the type dummies. The second lets the coefficient  $c$  on the reference payoff ratio be freely estimated (model DG2). The third model includes a set of controls for rival theories, to be explained in section 6 (model DG3). It can be readily verified that the parameter estimates are plausible and confirm AM’s results. The weakly selfish and weakly perfect substitutes-types have almost linear indifference curves (in DG1, the estimates are  $\hat{\rho} = .62$  and  $\hat{\rho} = .64$ , respectively). Weak egalitarians, on the other hand, have very strong curvature ( $\hat{\rho} = -4.43$  in DG1). At an equal allocation ( $x_{other}/x_{self} = 1$ ) the weakly selfish subjects value a marginal dollar to the other person as equivalent to a marginal 16 cents to themselves, while the weak perfect-substitutes types value a marginal dollar to the other at a full 77 cents to themselves. The weak egalitarians are not particularly altruistic when both subjects get the same amount, but the strong curvature means that the marginal rate of substitution changes dramatically as soon as one of the subjects gets less than the other. For ease of interpretation, and in order to facilitate comparison with the analysis of the CR data set

---

<sup>19</sup>Unless the subject is indifferent between all the available allocations. But then we would be very unlikely to observe corner choices in two out of three cases.

in the next section, table 2 converts the parameter estimates into marginal rates of substitution at representative allocations that put either the divider or the recipient behind (I report the MRS when  $x_{other}/x_{self} = .5$  and when  $x_{other}/x_{self} = 2$ ). Since the bulk of the sample consists of the choices of weakly selfish individuals, I shall mostly discuss the results for those subjects. The last panel of table 2, however, also reports the marginal rates of substitution and their set-dependence for weak egalitarians and weak perfect substitute-types.

Column 1 shows that all three weak types exhibit some aversion to inequality ( $\rho < 1$ ). In the model without set-dependence (DG1), weakly selfish dictators who get twice as much as the recipient value a dollar to the latter at 21 cents to themselves, but the MRS falls to .12 when the recipient gets twice as much as the dictator. For weak egalitarians, the inequality aversion is extreme, with a marginal cent to a disadvantaged recipient being worth 6.71 cents to the dictator, and a marginal cent to an advantaged recipient given zero value. The perfect substitutes-types have a consistently altruistic MRS, of .99 when they are ahead and .6 when they are behind.

Column 2 (tables 1 and 2) reports the estimates for the set-dependent model. It is clear that the reference payoff ratio matters a great deal for how generous the dictatorial divider chooses to be (the coefficient on  $r$  is highly significant, as can be seen in table 1). A weakly selfish divider values recipient payoffs more than three times as highly at the sample maximum reference payoff ratio ( $r = 2$ ) as at the sample minimum reference payoff ratio ( $r = .5$ ). When the divider gets twice as much as the recipient, the MRS ranges from .11 to .37; when she gets half as much, it ranges from .06 to .20. In contrast, the effect of changing the *actual* payoff ratio from  $x_{other}/x_{self} = .5$  to  $x_{other}/x_{self} = 2$  is “merely” a doubling of the marginal rate of substitution. These are economically significant magnitudes, and the evidence is that set-dependence is at least as important as set-independent distributive (or “outcome fairness”) concerns.

It is useful to have a sense of the “typical” effect of set-dependence. Table 2 gives one measure of the *average* effect and one measure of the *marginal* effect of set-dependence. The average effect is the change in the MRS (at sample averages) that results from a one standard deviation reduction in the reference payoff ratio (we look at a *reduction* in the other person’s relative fair claim in

order to be able to compare directly with an increase in his misbehaviour in CR's experiments). When the ratio of the recipient's reference payoff to the divider's reference payoff falls by one standard deviation ( $= .41$ ), the rate at which she is willing to transfer money to the recipient falls by  $-.07$  when she is ahead, and by  $-.04$  when she is behind.<sup>20</sup> The usefulness of this average measure, however, is limited for the purposes of generalisation or cross-study comparison (as we shall have occasion to perform in the next section). This is because we have no reason to think that the sample variation of the reference payoff ratio in this study is representative of any other experimental or non-experimental economic settings. Table 2 therefore also reports a marginal effect which is scaled to be comparable to other studies. The marginal measure is the elasticity of the marginal rate of substitution between own and other's payoff, with respect to the reference payoff ratio.<sup>21</sup> The table shows that the elasticity is close to unity: A one percent change in the reference payoff ratio is predicted to produce a .9 percent change in the MRS. For example, if the reference payoff ratio increased from .5 to .6, the marginal rate of substitution would be predicted to fall by 18%.

In sum, the econometrics suggests that set-dependence is an important phenomenon. The model statistics reject the restriction of the coefficient on the reference payoff ratio to zero. A likelihood ratio test shows that model DG2 improves on model DG1 in the dictator game data set at any conventional significance level. The effect of changes in the reference payoff ratio, moreover, is also important in weakly egalitarian individuals, as the second and third panels of table 2 show. Only for the one weak perfect substitute-type individual is the effect smaller (an elasticity of .2). This result makes intuitive sense. Perfect substitute-types care about efficiency; they are prepared to give up all the tokens when they are worth more to the recipient, but feel free to keep everything when the exchange rate favours themselves. It is natural to think that entitlements are a more important consideration for people who care about equity (like egalitarians), while if social efficiency is an agent's main goal, the "most fair" reference allocation is less relevant. It should be

---

<sup>20</sup>The fact that the absolute effect is larger when the agent is ahead than when she is behind is an artefact of the functional form. As long as  $\rho < 1$  the effect of a change in the altruism weight will be larger when the decision-maker is ahead than when she is behind.

<sup>21</sup>The formulas used to calculate both measures are derived in the appendix.

noted that the results for these other two weak types do not derive from separate estimations, but rather follow from the construction of the model. They do not, therefore, provide additional empirical evidence for the importance of set-dependence. But the intuitive plausibility of the results contributes to the overall appeal of the theory.

## **5 Revisiting Charness and Rabin (2002)**

### **5.1 Data and comparative framework**

The experiment described in the previous section shows that set-dependence occurs even in the absence of reciprocity. This section analyses a data set that allows us to compare the magnitude of the two effects. Gary Charness and Matthew Rabin (2002) report the results of an extensive series of experiments designed to test different hypotheses about other-regarding preferences against each other. The sample they use for calibration, and which I use below, contains data from 27 two-player games. Seven of these games were unilateral decision problems where one player chose between two allocations of payoffs to herself and another player. The remaining twenty games were sequential-move games, where the first mover decided between ending the game and implementing a certain allocation of payoffs, versus letting the second player decide between two other allocations of payoffs. Together, the dictatorial choices and the second-mover choices after “entry” by the first mover make up a data set of 903 observations of binary choices. This is the data set CR use to calibrate different distributive utility functions and assess which best captures behaviour across a wide range of games. Note that most of the variation in this sample is between subjects: No subject played more than four games, and many played only one. It is therefore not possible to assign the subjects to categories like in AM and in my experiment. When I compare my results from the previous section with estimates from the CR data set in subsection 5.3, I concentrate on the results for the weakly selfish subjects.

There are some dissimilarities between my analysis and the CR study that warrant discussion. They relate to the functional forms used by CR, who estimate linear utility functions on a data set of

discrete (binary) choices. They also estimate separate altruism parameters for their piecewise linear utility function depending on whether the decision-maker is ahead or behind the other person. Among the many utility functions they calibrate, the one that is relevant for this paper is:

$$U(\mathbf{x}) = \begin{cases} (1 - \alpha - \theta q)x_{self} + (\alpha + \theta q)x_{other} & \text{if } x_{other} \leq x_{self} \\ (1 - \beta - \theta q)x_{self} + (\beta + \theta q)x_{other} & \text{if } x_{other} > x_{self} \end{cases} \quad (8)$$

This model includes a reciprocity parameter  $\theta$  which measures the effect on the altruism weight (in a weighted average specification) of “misbehaviour” on the part of the other player. CR use an indicator variable  $q$  which is set to 1 if the other player has “misbehaved” and zero otherwise. Misbehaviour is defined as a first-mover choice which makes the second mover end up with a lower payoff than would otherwise have been possible. This reciprocity parameter is introduced completely *ad hoc*; the functional form in equation (8) is not theoretically derived.

CR chose this piecewise linear function in order to allow for the intuition that individuals are likely to put a different value on the other person’s payoff depending on their relative position. The dictator game experiment analysed in the previous section, on the other hand, presented subjects with continuous linear choice set, and the many interior choices that resulted are not well fitted with a linear utility function. In addition, the data set from the dictator game experiment has less variation in the case where the decision-maker is behind (most choice situations allowed the decision-maker to be both generous *and* keep a larger share for herself, which the majority of the subjects chose to do). I therefore did not estimate a separate altruism parameter for the two orthants of the payoff space. Nevertheless, the intuition that the degree of generosity is not the same when ahead and behind is captured with the curvature parameter  $\rho$ . As long as  $\rho < 1$ , the marginal rate of substitution of own for other’s payoff is monotonically decreasing in the payoff ratio  $x_{other}/x_{self}$ . In other words, the utility function (1) is a smooth approximation of CR’s original piecewise linear model without reciprocity (equation 8 with  $\theta$  restricted to zero).<sup>22</sup> I first

---

<sup>22</sup>This is true unless the two altruism weights in equation (??) have opposite signs, as assumed by “inequality aversion” models like Fehr and Schmidt (1999). Since CR do not find evidence of a negative weight on the other person’s payoff when the decision-maker is behind (they find  $\beta$  to be around zero), I maintain that the CES form is a good approximation to their functional form. The results below confirm that it does at least as good a job at describing

follow CR in analysing their data under the assumption that  $\theta = 0$  (no reciprocity), and solve the challenge of comparability by reproducing CR's results using equation 8, and repeating their analysis with my smooth approximation of their piecewise linear model (2 with  $c = 0$ ). I show below that it captures the same patterns as CR's original results and that by their own criteria, it in fact performs better than their piecewise linear function. I then let the marginal rate of substitution vary with the reference payoff ratio and compare the results to the calibration discussed in the previous section.

For the purpose of comparing the relative effects of reciprocity and set-dependence, we need to estimate a model that includes both effects. In order to remain faithful to CR's approach, we therefore complement utility function (1) with CR's own reciprocity term as a linear term in the weighted-average altruism parameter:

$$U_{\mathbf{r}}(\mathbf{x}) = \text{sign}(\rho) \left[ (1 - \alpha(\mathbf{r}) - \theta q) x_{self}^{\rho} + (\alpha(\mathbf{r}) + \theta q) x_{other}^{\rho} \right]. \quad (9)$$

That is, I let CR's own reciprocity term enter in exactly the same way as they. Recall that  $A \equiv \alpha / (1 - \alpha)$ , which means we can estimate exactly the same parameters as before since we have:

$$\alpha(\mathbf{r}) \equiv \frac{A(\mathbf{r})}{1 + A(\mathbf{r})} = \frac{a + cr}{1 + a + cr} \quad (10)$$

and the scalar  $r \equiv \frac{r_{other}}{r_{self}}$ .

This specification allows us to compare models with or without set-dependence and with or without reciprocity, by restricting  $c$  or  $\theta$  to be zero or be estimated freely.

Charness and Rabin use a logit estimation procedure to calibrate the different utility functions. Specifically, they find the parameter values that maximise the following likelihood function:

$$\Pr(\text{action A}) = \frac{e^{\gamma U(\text{action A})}}{e^{\gamma U(\text{action A})} + e^{\gamma U(\text{action B})}}. \quad (11)$$

---

the data.

$\gamma$  is a precision parameter — the higher is  $\gamma$ , the better the utility function predicts the choices. In addition, the likelihood value is an indicator of the relative performance of the different models. I use exactly the same procedure, substituting the utility function given in equation (9) as required. The estimated parameters in the most general model are  $\hat{a}$ ,  $\hat{\rho}$ ,  $\hat{c}$ ,  $\hat{\theta}$  and  $\hat{\gamma}$  ( $\hat{a}$ ,  $\hat{b}$ ,  $\hat{\theta}$  and  $\hat{\gamma}$  in the piecewise linear models). The findings of interest relate to the relative magnitudes of  $\hat{c}$  and  $\hat{\theta}$ .

## 5.2 Set-dependence versus reciprocity: Results

The parameter estimates are given in table 3.<sup>23</sup> For ease of interpretation and comparison with the dictator game experiment, table 4 restates all the results in terms of the marginal rate of substitution between payoff to self and payoff to the other person. With  $\rho \neq 1$ , the MRS depends on the payoff ratio. As before, table 4 therefore provides representative estimates for the MRS when ahead ( $x_{other}/x_{self} = .5$ ) and when behind ( $x_{other}/x_{self} = 2$ ). Column 1 reports the MRS for the original CR piecewise linear model. Column 2 gives the MRS for the smooth version of CR with neither reciprocity nor set-dependence. The third column includes set-dependence. The next three columns provide the equivalent estimates for the models with reciprocity: CR's original piecewise linear model with reciprocity, the smooth version with reciprocity, and the smooth version with reciprocity and set-dependence.

We first compare the estimates for the smooth approximation with those for CR's piecewise linear utility function. The smooth parametrisation, estimated in column 2, gives qualitatively similar estimates to the piecewise linear specification, which is of course to be expected since it summarises the same behaviour, just using a slope and a curvature parameter to characterise the indifference curve instead of two slope parameters. The main difference is that the disparity between the degree of altruism when ahead and behind is smaller with the curved utility function; the MRS changes less as we move from the advantaged to the disadvantaged position than it does in the piecewise linear formulation. This most likely reflects the rigidity of the piecewise linear form. In

---

<sup>23</sup>Note that table 3 reports the estimated weights  $\hat{a}$  and  $\hat{b}$  for a weighted sum specification. CR directly estimate the weights  $\hat{\alpha}$  and  $\hat{\beta}$  for a weighted average specification. Since  $\alpha \equiv a/(1+a)$  and  $\beta \equiv b/(1+b)$  in the absence of non-reciprocal set-dependence, it is straightforward to recover the desired weights. It can be verified that columns 1 and 4 of table 1 (models CR1 and CR3) accurately reproduce CR's results, given in table VI of their paper.

CR's specification, each of the slope parameters have to provide a linear fit to all the data points on one side of the 45-degree line. Having one slope parameter and one curvature parameter provides more flexibility. We observe that the precision parameter  $\gamma$  is an order of magnitude greater with the CES specification compared to the piecewise linear formulation. This should reassure us that the CES function does at least a good a job at capturing distributive preferences as does the CR's specification. All of these observations apply equally well in the case with reciprocity as in the case without (the comparisons are models CR2 versus CR1 and CR4 versus CR3, respectively).

Having reassured ourselves that the CES form adequately describes the CR results, we may further note the remarkable similarity of the estimated altruism in CR2 and in the dictator game experiment DG1, in both of which games preferences are constrained not to exhibit set-dependence. CR's subjects have an MRS of .26 when they are ahead and .13 when they are behind; this compares with the weakly selfish subjects in my experiment who have an MRS of .21 when they are ahead and .12 when they are behind. This should make us confident that the two data sets are comparable. The CES form does indeed seem to do a good job at describing other-regarding distributive preferences.

We now use the CES-RD utility function to investigate the role of set-dependence in CR data set and compare its effect the results from the dictator game experiment. Column 3 in tables 3 and 4 reports the results for the model that includes the reference payoff ratio (this is the same model as DG2 except for the WEAKEGAL and WEAKPSUB dummies). Table 3 shows that the coefficient on the reference payoff ratio is strongly statistically significant. In table 4 we can assess its quantitative importance. The effect on the MRS due to changes in the reference payoff ratio is remarkably strong. At the sample average reference payoff ratio, the MRS is close to its values in the restricted models (.24 when ahead, .12 when behind). But as it varies from its minimum to its maximum value in the sample, the utility function goes from being considerably spiteful — with a negative weight on the other person's payoff — to being quite a bit more generous than at the sample average. The estimates implies that an agent who has twice as much as the other player is willing to *give up* 41 cents to prevent the recipient from getting a marginal dollar when the

reference payoff ratio at its minimum value of .33, while at a reference payoff ratio of 1, she values a marginal dollar to the other player at 34 cents to herself. The average and marginal effects of set-dependence are between two and three times larger in the CR data set than the estimates for the dictator games. A one standard deviation fall in the reference payoff ratio leads to a .21 reduction in the MRS for an agent who is ahead (a .11 reduction for one who is behind). The elasticity of the MRS with respect to the reference payoff ratio is 4.3, more than four times as large as in the previous experiment. The next subsection addresses reasons for the disparity between the two data sets.

Column 4 reproduces CR's specification with reciprocity, and column 5 gives the results for the smooth approximation with reciprocity. The estimates reproduce CR's results, which show that reciprocity accounts for important variation in behaviour. It is clear, however, that reciprocity does not explain all the patterns of set-dependence. Comparing column 6 with column 3 shows that the effect of the reference payoffs is hardly diminished when we include Charness and Rabin's reciprocity parameter. The impact on the MRS of reference point changes in model SD2 is similar to that found in model SD1, and the likelihood ratio tests show that restricting the model from SD2 to CR4 to exclude set-dependence (by constraining  $c$  to be zero) is statistically rejected. Reciprocity also has an effect, however; altruism is uniformly weaker (and is uniformly negative) when the first mover has "misbehaved." A closer examination of CR's 27 games shows why both effects are important. The majority of the games either have no misbehaviour or a reference payoff ratio of 1. The non-reciprocal set-dependence results are driven by six games (one of which is a dictator-type game) that have  $r < 1$ , and the reciprocity results are caused by ten games in which the misbehaviour dummy is set to 1. It turns out that the two subsets of games only have one game in common. Unfortunately, the games are not sufficiently similar to make direct pairwise comparisons across particular games to get an intuitive sense of the relative magnitude two effects. As I have shown, however, the econometrics suggests that both reciprocal and non-reciprocal set-dependence plays a role in determining subjects' choices.

Which of the two effects is more important overall? Table 4 reports the average and the mar-

ginal effect of both a reduction in the reference payoff ratio and of misbehaviour by the first mover. We see in column 6 that when both effects are included, the average effect of first-mover misbehaviour, at  $-.24$ , is somewhat larger than that of non-reciprocal set-dependence at  $-.14$ .<sup>24</sup> But as I argued above, the average effects may not be suited for the purpose of comparing “typical” effects, since the opportunity for either reciprocal or non-reciprocal set-dependence to affect choice in the sample depends on how the choice situations are constructed. The larger average effect of reciprocity may simply reflect a data set with more variation in misbehaviour than in entitlements. However, it is possible to compare the marginal effect of the two phenomena, expressed as elasticities in order to sidestep the fact that the two variables are not measured on the same scale. Table 4 reports the point elasticity of the MRS with respect to the reference payoff ratio implied by the estimates of model SD2 to be 6.3. The point elasticity with respect to misbehaviour by the first mover is  $-2.0$  or  $-2.1$ , depending on whether set-dependence is directly included. Since misbehaviour is a zero/one-variable, however, it is not clear that the point elasticity is very meaningful. Table 4 therefore also reports the arc elasticity calculated between the extreme values (0 and 1) of the misbehaviour indicator, which is  $-3.5$  and  $-3.4$  in CR4 and SD2, respectively.<sup>25</sup> The elasticity measures are robust across the different models, and are essentially unaffected by whether only one or both of set-dependence and reciprocity are admitted in the model. The elasticities are of the same order of magnitude, but the non-reciprocal set-dependence elasticity is about twice as large (in absolute value) as the misbehaviour elasticity. This suggests that the dependence of fairness judgments on reference points in the payoff set is indeed an important part of distributive preferences, and at least as important as reciprocity.

### 5.3 Cross-study comparison

As mentioned above, comparing the first column of table 2 (model DG1) with the second column of table 4 (model CR2) reveals a remarkable robustness of the parameter estimates across the two

---

<sup>24</sup>For agents who are ahead; for disadvantaged agents the average effects on the MRS are  $-.13$  and  $-.08$ , respectively.

<sup>25</sup>The formulas used are given in the appendix. All the calculations are done at sample averages of  $r$  and  $q$ .

sets of experimental data. The behaviour of CR's subjects in stand-alone two-mover games, before considering reciprocity or set-dependence, is characterised by almost exactly the same utility function as the behaviour of weakly selfish players in my series of complex dictator games. The MRS of advantaged weakly selfish players in the dictator games is .21 and that of disadvantaged ones is .12; this compares with the .26 and .13 in the CR games. It is a very promising sign for research on other-regarding preferences that utility functions can be fitted with comparable parameter values for independent data sets. The fact that the observed behaviour is so similar in two such different studies should also give us more confidence in the results on set-dependence, which we summarise here:

We have fitted the same utility functions to two completely independent data sets; one consisting of choice behaviour from binary-choice, (mostly) two-move games, the other containing choice data from individual decision problems with convex choice sets. In both experiments, non-reciprocal set-dependence is shown to be a statistically and economically significant phenomenon (the results from SD1, SD2, and DG2) that conforms to theoretically derived predictions. While reciprocity motives do retain explanatory power in the CR data, they are quantitatively no more (and arguably less) important than non-reciprocal sensitivity to reference payoffs, which explain variations in behaviour even in unilateral decision problems where reciprocity does not have a role. I conclude that it is overly narrow to take reciprocity as the only explanation of why distributive preferences do not conform to the predictions of purely outcome-based models. The dichotomy should be between models which assume that only outcome fairness matters, versus models in which preferences over two given distributive outcomes may depend on non-outcome factors. The reciprocity motive is just one of many possible aspects of allocative processes that may determine the fairness of a resulting allocation, and should not be seen as *the* alternative to outcome-based models. The economic analysis of fairness should attempt to identify and characterise other such determinants and investigate their relative importance. This paper makes a step in that direction by demonstrating that a theoretically founded theory of non-reciprocal set-dependent reference payoffs can be empirically as successful as reciprocity theory.

In one respect only do the results on set-dependence differ in the two data sets. The effect is noticeably weaker in the dictator games than in the CR data (although it is still strong). There are three possible and mutually compatible reasons for this. One is that reciprocal and non-reciprocal set-dependence are sufficiently bound up with each other that the model estimated here cannot fully disentangle the two effects in the CR data set. Perhaps some of the strong effect of changes in the reference payoff ratio in that data set really reflects reciprocity that is not properly captured by Charness and Rabin's *ad hoc* misbehaviour variable. Even if this is true, the set-dependence effect in the dictator games can be taken as a lower bound on the "true" set-dependence in the CR data.

Another explanation is that the difference is driven by the discreteness of the choices in the CR experiment. When choice sets are continuous, any adjustment to changes in the reference payoff ratio can be fine-tuned; discrete choices, on the other hand, require discontinuous jumps. In the dictator games, an agent with convex preferences can choose a point which equalises the MRS with the price ratio. In the CR data agents must choose which is the most preferred of two alternatives, even though if they had the opportunity they would choose something in between the two options. This line of reasoning suggests that both non-reciprocal and reciprocal set-dependence effects may be overstated by the design of the CR experiments. As both of these problems (the unique focus on reciprocity and the use of discrete choice sets) are common characteristics of the experimental literature, future research should make an effort to address them directly.

A third factor is that in the dictator game experiment, subjects were confronted with a series of 36 decisions, whereas in the CR data, each subject only played between one and four games. More of the variation in the dictator game experiment is therefore within-subject variation. But if a subject thinks of the whole series of decisions as one problem, the fair reference allocation may not vary much from decision to decision. Instead, the subject may for example take as the reference allocation the most fair allocation in the whole series of games, not the most fair allocation available in each decision problem. This would cause the effect of set-dependence to be underestimated relative to the true effect that obtains in isolated, independent instances. Again, this suggests

that the dictator game results provide lower bounds on the magnitude of set-dependence. The true effect may well be closer to the larger estimates from the CR data.

## 6 Alternative theories: “Context-dependence”

As mentioned in the introduction, there are few developed accounts of how preferences over payoff allocations may be systematically affected by other factors, even though the reversal of distributive choices is a phenomenon that has been consistently observed in experiments. Reciprocity theory is the exception to that rule. Yet reversals have also been observed outside the context of preferences over payoff allocations and they have received more attention in those other contexts. In particular, studies in social psychology and marketing research have documented important effects of the set of alternatives on consumers’ and study subjects’ choice among various consumer products. Itamar Simonson and Amos Tversky call this phenomenon context-dependence, and understand by the “context” of an option either the set of concurrently available options, or the set of alternatives that is commonly associated with the item under consideration. Among the strong effects they have documented (Simonson and Tversky 1992, Tversky and Simonson 1993 report the studies) is what they call extremeness aversion — the tendency to move away from choices “at the edge” of the set of options. While Simonson and Tversky’s studies are mostly experimentally informed, they also offer a simple theoretical account of context-dependence in riskless choice, based on Kahneman and Tversky’s prospect theory of choice under uncertainty. In their theoretical framework, Simonson and Tversky assume that people do not assign an absolute value (“utility”) to options and then compare them, but that each option’s value is framed as a sum of functions of *advantages* and *disadvantages* relative to the alternatives. So if the choice objects have two valuable attributes — quality and cheapness, for example — then each object is evaluated according to how much more or less quality it has and how much cheaper or more expensive it is than each of the other available objects. If we assume that the value function is convex in the positive distance to other objects on each valuable dimension, then, as in prospect theory, “losses loom larger than gains,” or relative disadvantages loom larger than relative advantages of the same magnitude. This kind

of value function favours intermediate options over extreme ones, since a large advantage does not compensate as well for a large disadvantage as does a small advantage for a small disadvantage. So consumers will tend to choose an intermediate item rather than the best and most expensive or the worst and most inexpensive. This in turn means that adding higher-quality, higher-price items, or removing lower-quality, lower-price items from the choice set makes consumers more likely to choose higher-quality, higher-price items and *vice versa*.

Simonson and Tversky's context-dependence theory has not been applied to distributive preferences. But the reasoning behind it is readily transferable to that context, and the psychological mechanisms that underlie context-dependence in the space of consumer goods may well generate similar effects in the space of payoff allocations among individuals. Just as agents may evaluate a consumer object in terms of their advantages and disadvantages in terms of each valued attribute relative to other available objects, so they may evaluate a payoff allocation in terms of each individual's losses and gains relative to other available allocations. If context-dependence as theorised by Simonson and Tversky applies to distributive preferences, therefore, one should expect to find extremeness aversion in the data. If the set-dependence I documented in section 4 simply reflects such extremeness aversion, then the explanation in terms of reference payoffs should be forgotten in favour of a more mundane story that people are attracted to the middle of tradeoffs in general. To test whether set-dependence is independently important or whether it is just a reflection of Simonson-Tversky context-dependence, I therefore estimate a model that admits both extremeness aversion and set-dependence.

Note that extremeness aversion implies that *any* truncation of the budget set should induce agents to move their choice away from the truncation point and towards the middle of the new feasible set. Non-reciprocal set-dependence, on the other hand, predicts that only truncations that move the fair reference allocation should matter. From figure 3, we can see that each budget set is composed of up to four segments, and the decision problems vary according to which segments are part of the available set. The four segments consist of allocations that give the recipient more than  $2/3$  of the total monetary payoff ("outer high"), allocations that give the recipient between

$\frac{1}{2}$  and  $\frac{2}{3}$  of the total monetary payoff (“inner high”), allocations that give the recipient between  $\frac{1}{3}$  and  $\frac{1}{2}$  of the total monetary payoff (“inner low”) and allocations that give the recipient less than  $\frac{1}{3}$  of the total monetary payoff (“outer low”). According to context-dependence, removing these segments should induce people to choose points closer to the new interior of the available set. That is, removing a “low” segment should increase altruism, and removing a “high” segment should reduce it.<sup>26</sup> To control for Simonson-Tversky context-dependence, we therefore estimate model DG3, which adds to DG2 a set of four indicator variables whose value is set to one if the corresponding segment is *excluded* from the budget set, and zero otherwise. As before, the estimating equation is:

$$\frac{x_{other,ij}}{x_{self,ij} + x_{other,ij}} = \frac{A_{ij}^{\frac{1}{1-\rho_i}}}{p_j^{\frac{1}{1-\rho_i}} + A_{ij}^{\frac{1}{1-\rho_i}}} + \varepsilon_{ij}, \quad (12)$$

but now with

$$A_{ij} = a + cr_j + b_E WEAKEGAL_i + b_{PS} WEAKEPSUB_i + d_{OH} OUTERHIGH_j + d_{IH} INNERHIGH_j \quad (13)$$

$$+ d_{IL} INNERLOW_j + d_{OL} OUTERLOW_j, \quad (14)$$

where  $OUTERHIGH = 1$  if gifts above  $\frac{2}{3}$  of the pie are ruled out, *et cetera*. The coefficient  $c$  is still identified, since  $r$  is not a linear combination of the four controls. If set-dependence is just a reflection of Simonson and Tversky’s context-dependence, then the coefficient on  $r$  should diminish or become insignificant when we include the truncation controls. Those controls themselves should have statistically significant coefficients, since choices are sensitive to truncation from any side according to context-dependence. (Set-dependence, on the other hand, is only sensitive to

---

<sup>26</sup>This prediction is problematic in the one case where the two inner segments are removed and the outer two retained (the bottom-left-hand panel of figure 3). Simonson and Tversky consider the effects of manipulation the extremes of the choice set, rather than removing the middle. But since the one case which removes the interior of the choice set is symmetric (both the inner high and the inner low segments are removed), there is no reason to expect this to affect the pattern predicted in the main text. At most, this case may lead us to predict weaker downward (upward) effect on altruism of removing the inner high (low) segment than of removing the outer high (low) segment.

truncations that move the fair reference point, as already explained.)

Column 3 of tables 1 and 2 report the results. When the controls are included, the effect of the reference payoff ratio diminishes somewhat for disadvantaged agents, but increases considerably for dividers who are ahead of the recipients. The coefficient on  $r$  remains statistically significant. The parameter estimates imply that at the lowest reference payoff ratio in the sample, advantaged dividers now put negative value on payoffs to the recipient — one additional dollar to the recipient is equivalent to an 8 cent loss to the divider. At the highest reference payoff ratio in the sample, however, advantaged dividers are very generous, valuing a marginal dollar to the recipients at 53 cents to themselves. At the sample averages of the variables, the effect of a one standard deviation reduction in the reference payoff ratio is  $-.17$  when ahead ( $-.02$  when behind), and the elasticity of the MRS with respect to the ratio is 3.2, both for weakly selfish individuals. This compares with the DG2 estimates of  $-.07$  ( $-.04$  when behind) and  $.9$ , respectively. The increase in the magnitude of the estimate shows at the very least that the estimated set-dependence effect is robust to the inclusion of the truncation controls. For the other two weak types, we see that the estimates are virtually unchanged. Moreover, it is hard to find any evidence of extremeness aversion. The coefficients on the truncation indicators (reported in column 3 of table 1 with the raw parameter estimates) are all minute and statistically insignificant, except *INNERLOW*, and even that variable only affects the MRS by  $.08$  when it changes from zero to one (at egalitarian allocations where  $x_{other}/x_{self} = 1$ ). Subject behaviour in dictator games, we may conclude, exhibits clear set-dependence, and the preferences that rationalise this behaviour display sensitivity to reference payoff ratios and not extremeness aversion.

## 7 Conclusion

After starting to take seriously the social aspects of economic behaviour, economics has made strides in furthering our understanding of other-regarding preferences. The outcome-based models of distributive preferences are an important step in that progress, as is reciprocity theory and its insistence that preferences are not simply outcome-based. The model presented in section 2 sug-

gests how these models could be represented as special cases of an overall view of other-regarding motivations. Agents care not only about their own welfare, but also about fairness. Fairness in turn is determined by the distributive outcomes, but also by features of the process that influence notions of what individuals deserve — their “reference payoffs.” Reciprocity considerations may be an important part of how such reference payoffs are established, but they are not the only factor. In particular, reference payoffs may be *set-dependent* in a way that is completely independent of intentions.

The model of non-reciprocally set-dependent preferences employed in this paper nested earlier approaches such as outcome-based models and reciprocity models, as well as non-reciprocal set-dependence. This framework enabled me to test the different theories against two very different data sets. The results show a remarkable stability of parameter estimates across the data sets. This is a very satisfying finding, suggesting that the various *ad hoc* hypotheses about other-regarding preferences that have been proposed in the literature may be unified in a general theory that has strong empirical support. Further, the results suggest that while reciprocity is a factor, set-dependence is a separate and significant phenomenon. Which alternatives are present has a strong effect on what agents are thought to deserve, which in turn influences the behaviour of fairness-minded individuals. Reciprocity, therefore, is not the whole story. The feasible set matters through more than intentions.

## A Appendix: Derivations

### A.1 Derivation of estimating equation (5)

The first-order condition for maximising the utility function in equation (1) is:

$$\frac{x_{other}}{x_{self}} = \left( \frac{p}{\alpha(r) / (1 - \alpha(r))} \right)^{\frac{1}{\rho-1}} \quad (15)$$

where  $r = \frac{v_{other}}{v_{self}}$  and where  $p \equiv \frac{v_{self}}{v_{other}}$  is the price of giving (the ratio of the value of a token to oneself over the value to the other person). Expressing the gift as a ratio of the aggregate payoff, we get:

$$\frac{x_{other}}{x_{self} + x_{other}} = \frac{\left(\frac{\alpha(r)}{1-\alpha(r)}\right)^{\frac{1}{1-\rho}}}{p^{\frac{1}{1-\rho}} + \left(\frac{\alpha(r)}{1-\alpha(r)}\right)^{\frac{1}{1-\rho}}}. \quad (16)$$

So the estimating equation for the three models is:

$$\frac{x_{other,ij}}{x_{self,ij} + x_{other,ij}} = \frac{A_{ij}^{\frac{1}{1-\rho_i}}}{p_j^{\frac{1}{1-\rho_i}} + A_{ij}^{\frac{1}{1-\rho_i}}} + \varepsilon_{ij} \quad (17)$$

where  $\varepsilon_{ij}$  is a randomly distributed error for the observation of subject  $i$  in decision problem  $j$ . Note that  $A \equiv \alpha(r) / (1 - \alpha(r))$  is just the marginal rate of substitution along the 45 degree line ( $x_{other}/x_{self} = 1$ ) in the absence of reciprocity. Depending on the model we have different specifications of  $A$ :

1. In model DG1:

$$A_{ij} = a + b_E WEAKEGAL_i + b_{PS} WEAKEPSUB_i$$

2. In model DG2:

$$A_{ij} = a + cr_j + b_E WEAKEGAL_i + b_{PS} WEAKEPSUB_i$$

3. In model DG3:

$$A_{ij} = a + cr_j + b_E WEAKEGAL_i + b_{PS} WEAKEPSUB_i + d_{OH} OUTERHIGH_j \\ + d_{IH} INNERHIGH_j + d_{IL} INNERLOW_j + d_{OL} OUTERLOW_j.$$

In all models we have:

$$\rho_i = g + g_E WEAKEGAL_i + g_{PS} WEAKEPSUB_i.$$

## A.2 Calculating marginal rates of substitution elasticities

The MRS is:

$$MRS = \frac{\alpha(r) + \theta q}{1 - \alpha(r) - \theta q} \left( \frac{x_{other}}{x_{self}} \right)^{\rho-1}. \quad (18)$$

We continue the notation  $A \equiv \alpha(r) / (1 - \alpha(r))$  (so  $\alpha \equiv A / (1 + A)$ ) for the weight on the other person's payoff in a weighted sum formulation of the utility function when there is no reciprocity (cf. equations 1 and 5). The elasticity of the MRS with respect to the reference payoff ratio is

$$\frac{d \ln MRS}{d \ln r} = \frac{d \ln [\alpha(r) + \theta q]}{d \ln r} - \frac{d \ln [1 - \alpha(r) - \theta q]}{d \ln r} \quad (19)$$

$$= \frac{\partial \alpha(r)}{\partial r} \frac{r}{\alpha(r) + \theta q} + \frac{\partial \alpha(r)}{\partial r} \frac{r}{1 - \alpha(r) - \theta q} \quad (20)$$

$$= \frac{cr}{1 + A} \left[ \frac{1}{A + \theta q (1 + A)} + \frac{1}{1 - \theta q (1 + A)} \right] \quad (21)$$

$$= \left[ \frac{1 + \theta q}{A + \theta q (1 + A)} + \frac{\theta q}{1 - \theta q (1 + A)} \right] cr. \quad (22)$$

In the CR models,  $A = a + cr$ , so

$$\frac{d \ln MRS}{d \ln r} = \left[ \frac{1 + \theta q}{(a + cr)(1 + \theta q) + \theta q} + \frac{\theta q}{1 - \theta q (1 + a + cr)} \right] cr. \quad (23)$$

In the dictator games, there is no reciprocity, so

$$\frac{d \ln MRS}{d \ln r} = \frac{cr}{A}, \quad (24)$$

where  $A$  is as given in the previous subsection. In all cases, the elasticity is calculated at the sample averages of  $r$  and  $q$ .

In the CR models, the point elasticity of the MRS with respect to misbehaviour is

$$\begin{aligned} \frac{d \ln MRS}{d \ln q} &= \frac{d \ln [\alpha(r) + \theta q]}{d \ln q} - \frac{d \ln [1 - \alpha(r) - \theta q]}{d \ln q} \\ &= \frac{\theta q}{\alpha(r) + \theta q} + \frac{\theta q}{1 - \alpha(r) - \theta q} \end{aligned}$$

calculated at the sample averages of  $r$  and  $q$ . The arc elasticity of the MRS with respect to misbehaviour is:

$$\frac{\frac{MRS_{q=1} - MRS_{q=0}}{(MRS_{q=1} + MRS_{q=0})/2}}{\frac{1-0}{(1+0)/2}} = \frac{MRS_{q=1} - MRS_{q=0}}{(MRS_{q=1} + MRS_{q=0})} \quad (25)$$

$$= \frac{\frac{\alpha+\theta}{1-\alpha-\theta} - \frac{\alpha}{1-\alpha}}{\frac{\alpha+\theta}{1-\alpha-\theta} + \frac{\alpha}{1-\alpha}} \quad (26)$$

$$= \frac{(1-\alpha)(\alpha+\theta) - \alpha(1-\alpha-\theta)}{(1-\alpha)(\alpha+\theta) + \alpha(1-\alpha-\theta)} \quad (27)$$

$$= \frac{\theta}{\theta + 2(\alpha - \alpha^2 - \alpha\theta)} \quad (28)$$

where  $\alpha$  is calculated at the sample average of  $r$  in model SD2.

## B Appendix: Tables

Table 1: Censored tobit estimates for dictator games

<b>Model</b>	<b>DG1</b>	<b>DG2</b>	<b>DG3</b>
Set-dependence?	No	Yes	Yes
Robustness controls?	No	No	Yes
<b><u>Altruism weight estimates (A)</u></b>			
Constant (a)	0.16 (5.63)	0.02 (0.30)	-0.14 (4.05)
WEAKEGAL ( $b_E$ )	0.00 (0.02)	0.02 (0.15)	0.16 (1.42)
WEAKPSUB ( $b_{PS}$ )	0.61 (21.16)	0.63 (20.69)	0.74 (30.85)
Reference payoff ratio (c)		0.13 (2.97)	0.15 (3.45)
OUTERHIGH ( $d_{OH}$ )			-0.01 (0.58)
INNERHIGH ( $d_{IH}$ )			0.00 (0.15)
INNERLOW ( $d_{IL}$ )			0.08 (2.89)
OUTERLOW ( $d_{OL}$ )			0.01 (0.70)
Average effect of controls: (Control coefficients multiplied by sample average of control values)			0.04

Robust t-statistics, corrected for clustering by subject, in parentheses

**Table 1, continued**

<b>Model</b>	<b>DG1</b>	<b>DG2</b>	<b>DG3</b>
Set-dependence?	No	Yes	Yes
Robustness controls?	No	No	Yes
<b><u>Curvature estimates (<math>\rho</math>)</u></b>			
constant (g)	0.62 (5.25)	0.56 (5.45)	-0.43 (0.64)
WEAKEGAL ( $g_E$ )	-5.05 (1.81)	-5.01 (1.92)	-3.83 (1.99)
WEAKPSUB ( $g_{PS}$ )	0.01 (0.11)	0.06 (0.58)	1.03 (1.52)
<b><u>Implied parameter values, by type</u></b>			
<i>Altruism weight (A) by type, evaluated at sample averages:</i>			
Weakly selfish	0.16	0.15	0.05
Weakly egalitarian	0.16	0.17	0.21
Weakly perfect substitutes	0.77	0.77	0.79
<i>Curvature (<math>\rho</math>) by type, evaluated at sample averages</i>			
Weakly selfish	0.62	0.56	-0.43
Weakly egalitarian	-4.43	-4.45	-4.26
Weakly perfect substitutes	0.64	0.63	0.60
<b><u>Model statistics</u></b>			
sigma (s.d. of observation disturbance)	0.238	0.232	0.215
LL	-670.5	-658.0	-631.6
Chi-sq		24.98	52.98
p-value		0.000	0.000
Comparison model		DG1	DG2
n	2268	2268	2268

Robust t-statistics, corrected for clustering by subject, in parentheses

**Table 2: Marginal rates of substitution estimated from dictator games**

<b>Model</b>	<b>DG1</b>	<b>DG2</b>	<b>DG3</b>
Set-dependence?	No	Yes	Yes
Robustness controls?	No	No	Yes
<b>MARGINAL RATES OF SUBSTITUTION, WEAKLY SELFISH TYPES</b>			
<b><u>Altruism when ahead</u></b> (evaluated at $x_{\text{other}}/x_{\text{self}} = .5$ )	0.21		
r at sample average (r = 1)		0.20	0.12
r at sample minimum (r = .5)		0.11	-0.08
r at sample maximum (r = 2)		0.37	0.53
<b><u>Altruism when behind</u></b> (evaluated at $x_{\text{other}}/x_{\text{self}} = 2$ )	0.12		
r at sample average (r = 1)		0.11	0.02
r at sample minimum (r = .5)		0.06	-0.01
r at sample maximum (r = 2)		0.20	0.07
<b><u>Effect of fair claims ratio</u></b>			
<i>Average effect at sample averages</i>			
Effect on MRS of 1 s.d. change in r:			
- when ahead		-0.07	-0.17
- when behind		-0.04	-0.02
<i>Marginal effects at sample averages</i>			
Point elasticity of MRS with respect to r		0.9	3.2
<b><u>Model statistics</u></b>			
$\sigma$ (s.d. of observation disturbance)	0.238	0.232	0.215
LL	-670.5	-658.0	-631.6
n	2268	2268	2268
Likelihood-ratio tests:			
Chi-sq value		25.0	53.0
p-value		0.000	0.000
Comparison model		DG1	DG2

**Table 2, continued**

<b>Model</b>	<b>DG1</b>	<b>DG2</b>	<b>DG3</b>
Set-dependence?	No	Yes	Yes
Robustness controls?	No	No	Yes
<b>MARGINAL RATES OF SUBSTITUTION, WEAKLY EGALITARIAN TYPES</b>			
<b><u>Altruism when ahead</u></b> (evaluated at $x_{\text{other}}/x_{\text{self}} = .5$ )	6.71		
r at sample average (r = 1)		7.33	7.88
r at sample minimum (r = .5)		4.56	5.01
r at sample maximum (r = 2)		12.87	13.61
<b><u>Altruism when behind</u></b> (evaluated at $x_{\text{other}}/x_{\text{self}} = 2$ )	0.00		
r at sample average (r = 1)		0.00	0.01
r at sample minimum (r = .5)		0.00	0.00
r at sample maximum (r = 2)		0.01	0.01
<b><u>Effect of fair claims ratio</u></b>			
<i>Average effect at sample averages</i>			
Effect on MRS of 1 s.d. change in r:			
- when ahead		-2.27	-2.35
- when behind		0.00	0.00
<i>Marginal effects at sample averages</i>			
Point elasticity of MRS with respect to r		0.8	0.7
<b>MARGINAL RATES OF SUBSTITUTION, WEAK PERFECT SUBSTITUTES-TYPES</b>			
<b><u>Altruism when ahead</u></b> (evaluated at $x_{\text{other}}/x_{\text{self}} = .5$ )	0.99		
r at sample average (r = 1)		1.00	1.04
r at sample minimum (r = .5)		0.92	0.94
r at sample maximum (r = 2)		1.17	1.23
<b><u>Altruism when behind</u></b> (evaluated at $x_{\text{other}}/x_{\text{self}} = 2$ )	0.60		
r at sample average (r = 1)		0.60	0.60
r at sample minimum (r = .5)		0.55	0.54
r at sample maximum (r = 2)		0.69	0.71
<b><u>Effect of fair claims ratio</u></b>			
<i>Average effect at sample averages</i>			
Effect on MRS of 1 s.d. change in r:			
- when ahead		-0.07	-0.08
- when behind		-0.04	-0.05
<i>Marginal effects at sample averages</i>			
Point elasticity of MRS with respect to r		0.2	0.2

**Table 3: Logit estimates for Charness and Rabin's (2002) data set**

<b>Model</b>	<b>CR1</b>	<b>CR2</b>	<b>SD1</b>	<b>CR3</b>	<b>CR4</b>	<b>SD2</b>
Reciprocity included?	No	No	No	Yes	Yes	Yes
Set-dependence included?	No	No	Yes	No	No	Yes
<b><u>Piecewise linear specification</u></b>						
Altruism weight when ahead (a)	0.73 (11.85)			0.73 (12.38)		
Altruism weight when behind (b)	-0.01 (0.75)			0.02 (1.14)		
Reciprocity ( $\theta$ )				-0.11 (3.31)		
<b><u>Smooth specification</u></b>						
Altruism weight (a)		0.18 (3.96)	-0.57 (3.74)		0.27 (6.37)	-0.42 (3.13)
Reference payoff ratio (c)			0.82 (4.76)			0.76 (4.98)
Curvature ( $\rho$ )		0.51 (8.41)	0.52 (9.09)		0.56 (10.55)	0.59 (11.59)
Reciprocity ( $\theta$ )					-0.38 (4.97)	-0.38 (4.61)
<b><u>Model statistics</u></b>						
$\gamma$	0.014	0.175	0.166	0.015	0.174	0.149
LL	-527.7	-558.6	-551.2	-523.1	-550.1	-541.2
n	903	903	903	903	903	903
Likelihood-ratio tests:						
Chi-sq value			14.9		17.0	17.9
p-value			0.000		0.000	0.000
Comparison model			CR2		CR2	CR4

Robust t-statistics in parentheses

**Table 4: Marginal rates of substitution estimated from Charness and Rabin's (2002) data set**

<b>Model</b>	<b>CR1</b>	<b>CR2</b>	<b>SD1</b>	<b>CR3</b>	<b>CR4</b>	<b>SD2</b>
Reciprocity included?	No	No	No	Yes	Yes	Yes
Set-dependence included?	No	No	Yes	No	No	Yes
<b><u>Altruism when ahead</u></b> ( $x_{\text{other}}/x_{\text{self}} = .5$ )						
<i>Other has not misbehaved</i> ( $q = 0$ ):	0.73	0.26		0.73	0.36	
r at sample average ( $r = 0.91$ )			0.24			0.36
r at sample minimum ( $r = 0.33$ )			-0.41			-0.23
r at sample maximum ( $r = 1$ )			0.34			0.45
<i>Other has misbehaved</i> ( $q = 1$ ):	0.73	0.26		0.45	-0.20	
r at sample average ( $r = 0.91$ )			0.24			-0.20
r at sample minimum ( $r = 0.33$ )			-0.41			-0.49
r at sample maximum ( $r = 1$ )			0.34			-0.16
<b><u>Altruism when behind</u></b> ( $x_{\text{other}}/x_{\text{self}} = 2$ )						
<i>Other has not misbehaved</i> ( $q = 0$ ):	-0.01	0.13		0.02	0.20	
r at sample average ( $r = 0.91$ )			0.12			0.20
r at sample minimum ( $r = 0.33$ )			-0.21			-0.13
r at sample maximum ( $r = 1$ )			0.18			0.25
<i>Other has misbehaved</i> ( $q = 1$ ):	-0.01	0.13		-0.08	-0.11	
r at sample average ( $r = 0.91$ )			0.12			-0.11
r at sample minimum ( $r = 0.33$ )			-0.21			-0.28
r at sample maximum ( $r = 1$ )			0.18			-0.09
<b>Effect of reciprocity and fair claims ratio</b>						
<i>Average effects at sample averages</i>						
Effect on MRS of 1 s.d. change in r:						
- when ahead			-0.21			-0.14
- when behind			-0.11			-0.08
Effect on MRS of 1 s.d. change in q:						
- when ahead					-0.24	-0.24
- when behind					-0.13	-0.13
<i>Marginal effects at sample averages</i>						
Point elasticity of MRS with respect to r			4.3			6.3
Point elasticity of MRS with respect to q					-2.1	-2.0
Arc elasticity of MRS with respect to q					-3.5	-3.4

## References

- Andreoni, James and John Miller. 2002. "Giving According to Garp: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70(2):737–753.
- Andreoni, James, Paul M. Brown and Lise Vesterlund. 2002. "What Makes an Allocation Fair? Some Experimental Evidence." *Games and Economic Behavior* 40:1–24.
- Andreoni, James and Ragan Petrie. 2004. "Beauty, Gender and Stereotypes: Evidence From Laboratory Experiments." Unpublished manuscript.
- Blount, Sally. 1995. "When Social Outcomes Aren't Fair: The Effect of Causal Attribution on Preferences." *Organizational Behavior and Human Decision Processes* 63:131–144.
- Bolton, Gary E and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *The American Economic Review* 90(1):166–193.
- Bolton, Gary E, Jordi Brandts and Alex Ockenfels. 1998. "Measuring Motivations in the Reciprocal Responses Observed in a Dilemma Game." *Experimental Economics* 1:207–219.
- Brandts, Jordi and Gary Charness. 1999. "Retribution in a Cheap-Talk Experiment." Universitat Pompeu Fabra Working Paper.
- Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. The Roundtable Series in Behavioral Economics. New York and Princeton: Princeton University Press and Russell Sage Foundation.
- Charness, Gary and Matthew Rabin. 2002. "Understanding Social Preferences With Simple Tests." *The Quarterly Journal of Economics* 117:817–869.
- Cox, James C. and Daniel Friedman. 2002. "A Tractable Model of Reciprocity and Fairness." Unpublished manuscript.

- Dufwenberg, Martin and Georg Kirchsteiger. 1998. "A Theory of Sequential Reciprocity." Tilburg Center for Economic Research Discussion Paper No. 9837.
- Elster, Jon. 1983. *Sour Grapes: Studies in the subversion of rationality*. Cambridge, UK and Paris: Maison des Sciences de l'Homme and Cambridge University Press.
- Falk, Armin, Ernst Fehr and Urs Fischbacher. 1999. "On the Nature of Fair Behavior." University of Zürich Working Paper No. 17.
- Falk, Armin and Urs Fischbacher. 2000. "A Theory of Reciprocity." University of Zürich Working Paper No. 6.
- Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition and Cooperation." *The Quarterly Journal of Economics* 114:817–868.
- Fehr, Ernst and Simon Gächter. 2002. "Altruistic punishment in humans." *Nature* 415(10 January 2002):137–140.
- Fischbacher, Urs. 1999. "z-Tree: Zurich Toolbox for Readymade Economic Experiments. Experimenter's Manual." University of Zürich Working Paper No. 21.
- Gächter, Simon and Arno Riedl. 2002. "Moral Property Rights in Bargaining." University of Zürich Working Paper No. 113.
- Güth, Werner, Steffen Huck and Wieland Müller. 2001. "The Relevance of Equal Splits in Ultimatum Games." *Games and Economic Behavior* 37:161–169.
- Kahneman, Daniel, Jack L. Knetsch and Richard H. Thaler. 1986a. "Fairness and the Assumptions of Economics." *Journal of Business* 59(4):S285–S300.
- Kahneman, Daniel, Jack L. Knetsch and Richard H. Thaler. 1986b. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *The American Economic Review* 76(4):728–41.

- Prasnikar, Vesna and Alvin E. Roth. 1992. "Considerations of Fairness and Strategy: Experimental Data from Sequential Games." *The Quarterly Journal of Economics* 107:865–888.
- Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *The American Economic Review* 83(5):1283–1302.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Roth, Alvin E. 1995. Bargaining Experiments. In *Handbook of Experimental Economics*, ed. J Kagel and Alvin E. Roth. Princeton, NJ: Princeton University Press pp. 253–348.
- Sandbu, Martin Eiliv. 2003. "Axiomatic Foundations for Reference-Dependent Distributive Preferences." Chapter 2 of Ph.D. dissertation, Harvard University.
- Sen, Amartya. 1997. "Maximization and the Act of Choice." *Econometrica* 65(4):745–779.
- Simonson, Itamar and Amos Tversky. 1992. "Choice in Context: Tradeoff Contrast and Extreme-ness Aversion." *Journal of Marketing Research* XXIX:281–295.
- Tversky, Amos and Itamar Simonson. 1993. "Context-dependent Preferences." *Management Science* 39(10):1179–1189.