Reference Dependent Altruism

Breitmoser, Yves and Tan, Jonathan H.W.

7 January 2014

Reference Dependent Altruism

Yves Breitmoser
HU Berlin

Jonathan H. W. Tan[*]
University of Nottingham

January 7, 2014

Abstract

In view of behavioral patterns left unorganized by current social preference theories, we propose a theory of reference dependent altruism (RDA). With RDA, one's degree of altruism increases at reference points. It induces equity and efficiency effects that are conditional on whether or not payoffs meet reference points. We verify the theory first by experimentally analyzing majority bargaining, where observed behavior contradicts existing theories but confirms RDA. Using parameter estimates from majority bargaining, we then make out-of-sample predictions for Charness-Rabin, Engelmann-Strobel, and Bolton-Ockenfels games. RDA organizes these seemingly disparate games out-of-sample, which validates our hypothesis that pro-social behavior primarily relates to reference points.

JEL–Codes: C72, C78, D72

Keywords: bargaining, non-cooperative game, laboratory experiment, social preferences, quantal response equilibrium

[*] We gratefully acknowledge comments of Friedel Bolle, Maria Montero, and participants of the ESA World Meeting 2013, and financial funding of the DFG (projects BO 747/10–1 and BR 4648/1–1). Corresponding author: Jonathan Tan. Address: Nottingham University Business School, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, United Kingdom, email: *jonathan.tan@nottingham.ac.uk*, Phone/Fax: +44 115 846 6414/6667.

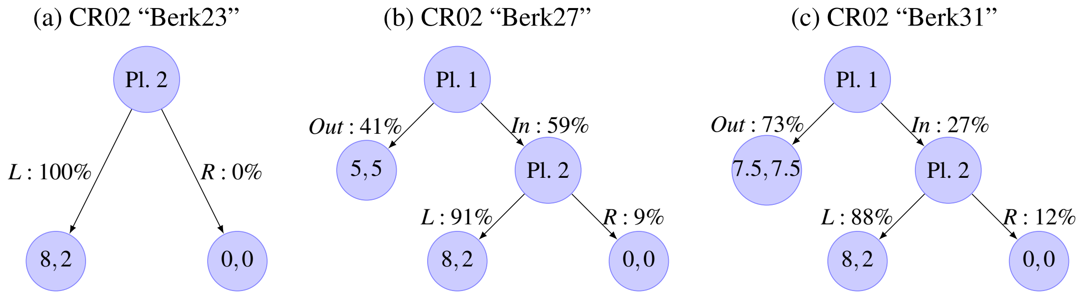
1 Introduction

Research on social preferences took a progressive leap with theories of inequity aversion proposed by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). They helped us make sense of deviations from egoistic behavior in ultimatum, dictator, market, and trust games—observations previously considered "anomalous" (Thaler, 1988; Camerer and Thaler, 1995). Subsequent work of Charness and Rabin (2002), Engelmann and Strobel (2004), and Kritikos and Bolle (2001) showed that subjects were willing to give up equity for efficiency and reciprocity. This suggests an interplay of egoistic, equity, efficiency, and reciprocity concerns, but with respective weights that vary across games. For instance, Bolton and Ockenfels (2006) argue that there is a "trade-off between efficiency and equity motives" (p. 1906) and that "perceptions of fairness depend on context" (p. 1909). Applying a model of social preferences with context dependent weights to economic analysis is obviously problematic.

To solve this problem, our paper proposes and tests a model of *reference dependent altruism* (RDA). Utilities are simple linear functions $u_i = x_i + \alpha x_j$ of payoffs (x_i, x_j), but the altruism weight α depends on the relation of the payoff x_i to a reference point. We estimate the model parameters on novel experimental data on majority bargaining, which we show to be particular suitable to discrimate social preference theories, and then use these estimates to make out-of-sample predictions for three widely discussed data sets. RDA explains behavior best both in-sample and out-of-sample. Our estimates indicate that 45% of the subjects use their ex-ante expected payoff as reference points, while 55% of the subjects use the opponents' payoffs as reference points. All subjects are highly efficiency concerned when they are above their reference points and only mildly altruistic below their reference point. The conjunction of the utility jump and the altruism weight change at the reference point captures both efficiency and equity concerns, as well as reciprocity in the sense of Rabin (1993).

To illustrate, Figure 1 shows three games analyzed by Charness and Rabin (2002, CR02). These games are mini dictator and ultimatum games testing whether player 2 prefers the allocation $(x_1, x_2) = (8, 2)$ to $(0, 0)$, and whether the preference depends on the payoff allocation in an outside option foregone by player 1. Choices in Berk23 are similar to those faced by proposers in mini dictator games, and choices in Berk27 and Berk31 are similar to those faced by responders who have received inequitable offers in mini ultimatum games. If player 2 is inequity averse in the sense of Fehr

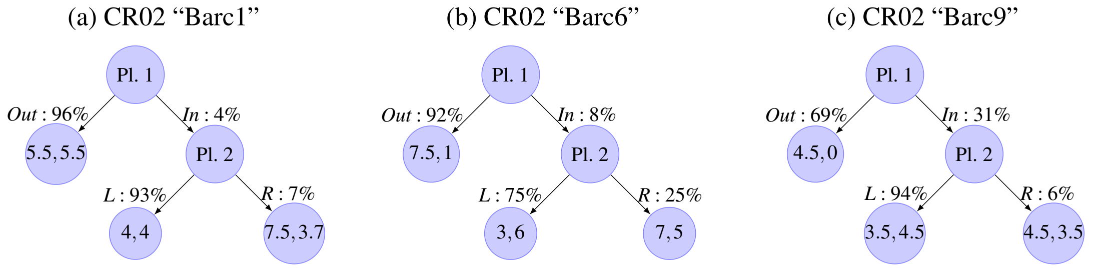
Figure 1: Example games (with percentages of choices) from Charness and Rabin (2002, CR02) where egoism or efficiency dominates inequity aversion


(a) CR02 "Berk23"

(b) CR02 "Berk27"

(c) CR02 "Berk31"

Figure 2: Example games (with percentages of choices) from Charness and Rabin (2002) where egoism or inequity aversion dominates efficiency or reciprocity


(a) CR02 "Barc1"

(b) CR02 "Barc6"

(c) CR02 "Barc9"

and Schmidt (1999), henceforth abbreviated as "FS", with standard envy weights (see e.g. Fehr and Schmidt, 2010), she would pick $(0,0)$ in all cases. In the experiment, however, all subjects in Berk23 and around nine of ten subjects in Berk27 and Berk31 chose the egoistic and efficient allocation $(8,2)$. In these cases, efficiency concerns therefore outweigh equity concerns.

In contrast, consider three other games by Charness and Rabin in Figure 2. These games resemble mini trust games. Player 1's decision to enter the game unlocks the efficient allocation R, and by picking R player 2 can reciprocate player 1. Yet, in all three games, most players 2 picked the egoistic or equitable option L. That is, equity concerns outweigh efficiency (and reciprocity) now—although the personal costs of player 2 are very small (e.g. in Barc1) in relation to the efficiency gain that would follow from picking R.[1]

[1] Note that egoism, i.e. maximization of pecuniary payoffs, explains behavior in these examples, which where chosen to highlight the tension between efficiency and equity. Egoism is not a principle organizing behavior in general, though, as has been shown in the literature and will be ascertained in our out-of-sample analysis in Section 6.

Table 1: Experimental games (with percentages of choices) from Engelmann and Strobel (2004, ES04) and Bolton and Ockenfels (2006, BO06) where efficiency concerns dominate inequity aversion and vice versa, respectively

(a) ES04 "Envy-Ny"

	Pl. 2 chooses		
Allocation	*A*	*B*	*C*
Player 1	16	13	10
Player 2	7	8	9
Player 3	5	3	1
Subj. choices	76.7%	13.3%	10%

(b) ES04 "Envy-N"

	Pl. 2 chooses		
Allocation	*A*	*B*	*C*
Player 1	16	13	10
Player 2	8	8	8
Player 3	5	3	1
Subj. choices	70%	26.7%	3.3%

(c) ES04 "RPG-P"

	Pl. 2 chooses		
Allocation	*A*	*B*	*C*
Player 1	14	11	8
Player 2	4	4	4
Player 3	5	6	7
Subj. choices	60%	6.7%	33.3%

(d) BO06 "Games I, II and III"

	Pl. 2 chooses			
Allocation	A	B_I	B_{II}	B_{III}
Player 1	13	19	27	27
Player 2	13	13	1	9
Player 3	13	13	17	9
Subj. choices	-	83.3%	8.3%	14.6%

Note: In the games of ES04, inequity aversion predicts C, while efficiency concerns predict A. In Games I, II and III of BO06, subjects choose between A and B_I, A and B_{II}, or A and B_{III}, respectively (e.g. 83.3% choose B_I over A); inequity aversion predicts A, while efficiency concerns predicts B_x. The observations choice frequencies reported for BO06 are for their "equal opportunities mode", which has a random role allocation procedure similar to that used by Engelmann and Strobel (2004).

Related results on subjects switching between efficiency and equity concerns are presented by Engelmann and Strobel (2004, ES04) and Bolton and Ockenfels (2006, BO06). In Engelmann and Strobel's dictator games shown in Tables 1a–1c, most subjects (as player 2) chose the efficient option A over the equitable option C, despite the latter being the unequivocal prediction of FS inequity aversion for all possible parameter values. Compare this to the majority rule voting games of Bolton and Ockenfels shown in Table 1d. In Game I, most subjects (as player 2) picked the efficient option B_I over the equitable option A, contradicting theories of inequity aversion again, but in Games II and III where efficiency gains in B require larger personal losses relative to A (for player 2), most subjects chose equity (A) over efficiency (B).

With these mixed results, one may be tempted to conclude that egoism, efficiency and equity concerns all shape behavior, but the empirical weights that one would estimate under this assumption would be unstable across games (Bolton and Ockenfels, 2006). We argue that the pattern along which these observations can be organized follows the relation of the players' payoffs to two simple reference points. Our basic idea is captured in a model of *reference dependent altruism* (RDA) which posits that one's altruism weight changes at one's reference point.

Reference points may be one's ex-ante expectation or the opponent's payoff. We refer to them as *absolute* and *relative* reference points, respectively, and estimate that the majority of subjects (55%) uses the opponents' payoffs as reference payoffs. This explains, for example, why subjects seem to value equity over efficiency in CR02's trust games (Figure 2) and in BO06's voting games *II* and *III* (Table 1d), while they value efficiency over equity in CR02's mini-ultimatum games (Figure 2) and in ES04's three-player dictator games. In the latter cases, the players' relation to their reference points is constant across options, while in the former cases, their payoff is above the reference point in one option and below it in the other. Now, if the utility is $u_i = x_i + 0.9x_j$ if $x_i \geq x_j$, i.e. above the reference point, and $u_i = x_i + 0.2x_j$ if $x_i < x_j$, i.e. below it, then reaching the reference point implies a utility jump on the order of $0.7x_j$.[2] This simultaneously explains why player 2 prefers $(4,4)$ to $(7.5,3.7)$ in Barc1 and $(8,2)$ to $(0,0)$ independently of outside options in Figure 1, while also explaining behavior in standard dictator and trust games.

Our test of RDA is twofold. First, we identify reference dependence of altruism in majority bargaining following Baron and Ferejohn (1989): The game proceeds in rounds, in each round one player is randomly assigned to propose an allocation, and the proposal is implemented if a majority votes in the affirmative (otherwise a new round begins). Majority bargaining is particularly interesting for two reasons. On the one hand, comparing the indefinite horizon game with a one round variant allows us to sharply separate CES altruism (Andreoni and Miller, 2002), FS inequity aversion (Fehr and Schmidt, 1999), CR reciprocity (Charness and Rabin, 2002), and RDA with either absolute or relative reference points. On the other hand, behavior in the canonical indefinite horizon multi-player bargaining game (Fréchette et al., 2005a,b; Fréchette et al., 2012; Montero et al., 2008) contradicts current theories of social preferences (Montero, 2007). Understanding preferences in this context is important, as it facilitates analyses of majority decisions under different institutional designs, e.g. in boards

[2] Utility jumps at reference points have previously been evidenced in the context of demand bargaining (Breitmoser and Tan, 2013) and resemble recently detected utility jumps in risk and time preferences (Diecidue and Van de Ven, 2008; Andreoni and Sprenger, 2012a,b).

and assemblies (Snyder Jr et al., 2005). Due to the utility jump at the reference point, RDA predicts that subjects take risks to reach the reference point, rejecting proposals below their reference point in the hope of reaching it in the next round. This prediction explains behavior, and indeed, RDA fits the experimental data in both games better than existing theories—qualitatively (Section 4) and quantitatively (Section 5).

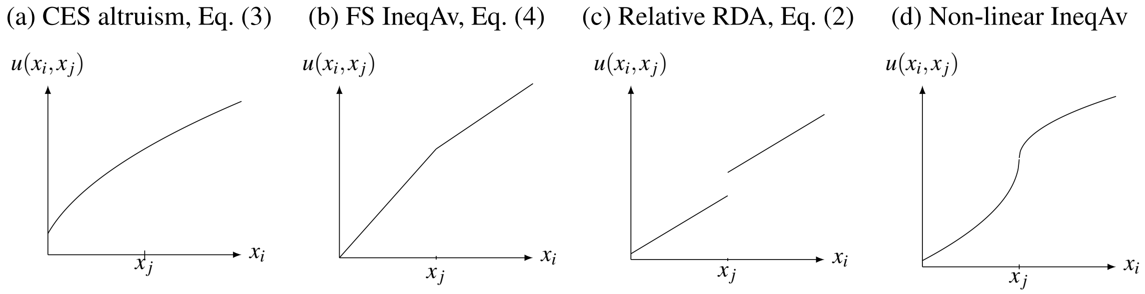
Second, we show that RDA not only fits better, but also predicts better out-of-sample. We test the validity of RDA in a wide domain of games well known in the literature as tests of social preferences. In Section 6, we use RDA with parameter estimates from our experiment to make out-of-sample predictions for Charness-Rabin dictator and response games, Engelmann-Strobel dictator games, and Bolton-Ockenfels voting games. We compare the accuracy of RDA's predictions against those of models by Fehr and Schmidt (1999) and Charness and Rabin (2002). The evidence supports RDA as a general theory that is consistently valid across this spectrum of games. Limitations and future research are discussed in the concluding Section 7.

2 Reference dependent altruism

The relevance of reference points in decision making has been established in a large variety of contexts. The best-known example is the Prospect theory of Kahneman and Tversky (1979) for individual choice under risk. Reference points also shape market interactions with risk (Kőszegi and Rabin, 2006, 2007) and even strategic choice as in loss-aversion equilibria (Shalev, 2002). A related branch of literature established the relevance of reference points with respect to social preferences. Examples include evidence on focal points in alternating bargaining (Murnighan et al., 1988), social comparisons to reference points in three-player ultimatum bargaining (Knez and Camerer, 1995), and norms as reference points in dictator and ultimatum games (Andreoni and Bernheim, 2009; Bicchieri and Chavez, 2010).

Our model of reference dependent altruism (RDA) builds on these observations establishing the relevance of reference points. RDA-players have simple altruistic utility functions, such as $u_i = x_i + \alpha x_j$ with x_i, x_j being the players' payoffs, where the altruism weight α depends on the relation of x_i to a reference point. That is, *utility parameters* are reference dependent. Players may adopt various, possibly even multiple reference points, but in the following we focus on two simple and widely discussed ref-

Figure 3: The various utility functions for $\alpha = 0.5$ and $\beta = 0.1$


(a) CES altruism, Eq. (3) (b) FS IneqAv, Eq. (4) (c) Relative RDA, Eq. (2) (d) Non-linear IneqAv

Note: In this case, where the opponent's payoff is given and he cannot "misbehave", CR reciprocity model is equivalent to FS inequity aversion, and the branches of RRDA have the same slope (while they have different slopes along the Pareto-frontier, $x_i + x_j = C$, in bargaining games).

erence points: the *ex-ante expected payoff* (following Kőszegi and Rabin, 2006, 2007) and the *opponent's payoff* (following Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). We will refer to the former as an absolute reference point and to the latter as a relative reference point.[3] The utilities of players with absolute or relative reference dependent altruism, abbreviated as ARDA and RRDA, respectively, are

$$U_i^{ARDA}(\mathbf{x}) = x_i + \alpha \sum_{j \neq i} x_j \cdot I_{x_i \geq x_i^*} + \beta \sum_{j \neq i} x_j \cdot I_{x_j < x_i^*} \tag{1}$$

$$U_i^{RRDA}(\mathbf{x}) = x_i - \alpha \sum_{j \neq i} x_j \cdot I_{x_i < x_j} - \beta \sum_{j \neq i} x_j \cdot I_{x_i \geq x_j}, \tag{2}$$

for all payoff profiles $\mathbf{x} \in \mathbb{R}^n$. Here, x_i^* is the absolute reference point of the ARDA player. We define x_i^* to be the ex-ante equilibrium payoff of i in case all players are payoff maximizers (assuming this value is unique). Thus, $\alpha > \beta$ implies that an ARDA-player becomes (more) altruistic toward all his opponents once his payoff exceeds the ex-ante expectation. In contrast, $\alpha > \beta$ implies for RRDA-players that they become (more) altruistic toward a specific opponent once his payoff exceeds that of this particular opponent.

The difference between RDA and other utility functions, e.g. those of Kőszegi and Rabin (2006, 2007), Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), is that RDA implies a utility jump the reference point (if the opponents' payoffs are not zero). ARDA with $\alpha > \beta$ imply that utility jumps by $(\alpha - \beta) \sum_{j \neq i} x_j$ once x_i crosses

[3] Additional reference points such as effort and production may become relevant if the players share the output of work (Cherry et al., 2002; Cappelen et al., 2007).

the reference point x_i^*. RRDA with $\alpha > \beta$ imply that utility jumps by $(\alpha - \beta)x_j$ when i overtakes j. These utility jumps are the key difference to existing theories of social preferences,[4] most notably CES altruism and FS inequity aversion, which are respectively defined as

$$U_i^{CES}(\mathbf{x}) = \Big((1-\alpha)\cdot(1+x_i)^{\beta} + \alpha/n \textstyle\sum_{j\neq i}(1+x_j)^{\beta}\Big)/\beta, \tag{3}$$

$$U_i^{FS}(\mathbf{x}) = x_i - \sum_{j\neq i} \alpha\,(x_j - x_i)\cdot I_{x_i<x_j} - \sum_{j\neq i} \beta\,(x_i - x_j)\cdot I_{x_i\geq x_j}. \tag{4}$$

CES utility functions are used frequently in analyses of dictator and public goods games (Andreoni and Miller, 2002; Goeree et al., 2002; Cappelen et al., 2007). The functional form used here has been proposed by Cox et al. (2007) and is adopted for its numerical stability. Here, α measures the degree of altruism, and $1/(1-\beta)$ is the elasticity of substitution. FS inequity aversion captures behavior in ultimatum and trust games, amongst others (Fehr and Schmidt, 1999, 2010). Generally, it is assumed that envy (α) outweighs guilt (β) and that guilt is bounded, i.e. $0 \leq \beta < \alpha$ and $\beta < 1/2$. Charness and Rabin (2002) extend FS inequity aversion by introducing a reciprocal component with weight θ. In two-player cases, CR utilities are

$$U_i^{CR}(\mathbf{x}) = (1 - \rho\cdot r - \sigma\cdot s - \theta\cdot q)x_i + (\rho\cdot r + \sigma\cdot s + \theta\cdot q)x_j, \tag{5}$$

where $r = 1$ if $x_i \geq x_j$, else $r = 0$, $s = 1$ if $x_i < x_j$, else $s = 0$, and $q = -1$ if j previously "misbehaved" by making a welfare-reducing decision, else $q = 0$. CR reciprocity reduces to FS inequity aversion for $\theta = 0$ (after relabeling parameters $\beta = \rho$ and $\alpha = -\sigma$). If $\theta > 0$, however, CR-players tend to punish welfare-diminishing behavior of their opponents, as their altruism weights decrease after such "misbehavior".

The two RDA utility functions, FS inequity aversion, and CR reciprocity are all instances of reference dependent preferences in the sense of Neilson (2006), and the technical differences are subtle. RRDA and FS inequity aversion model players comparing their own payoff to their opponent's payoff, U^{ARDA} models players comparing their payoff to their ex-ante expectation, and U^{CR} model players comparing both, their own payoff to their opponent's payoff and the aggregate welfare to the welfare maximum. Thus, CR reciprocity assumes two reference points, and U^{CR} is continuous at

[4] Similar jump discontinuities have been observed recently in risk and time preferences (Diecidue and Van de Ven, 2008; Andreoni and Sprenger, 2012a,b).

the opponent's payoff but exhibits a jump discontinuity at the welfare maximum. U^{FS} is continuous on the whole domain, while both RDA utility functions exhibit jump discontinuities at their respective reference points.

The key implication of a utility jump is that the player is willing to take risks to reach the reference point. This relates RDA to *S*-shaped utility functions as in Prospect theory, see e.g. Figure 3d. The difference is that RDA players take risks *only* to reach the reference point. They are risk neutral as long as they are securely above or below the reference point. Thus, the only difference between RRDA and FS inequity aversion is indeed the utility jump at the reference point, and the only difference between RRDA and ARDA, in turn, is the location of the reference point—which allows us to specifically analyze utility continuity. An additional analysis of the curvature along the utility branches requires further experiments.

3 The experimental games

Our experiment implements majority bargaining as modeled by Baron and Ferejohn (1989). Their random-proposer model is the canonical model of decision making in committees and parliaments with empirical applications ranging from estimating proposer power in the US Congress (Knight, 2005) to modeling labor relations Okada (2011). The Baron-Ferejohn game proceeds in rounds, with indefinite time horizon. In each round, a player is randomly recognized as proposer, he proposes an allocation (of the "Dollar"), and the other players vote on it. The proposal is implemented if a majority votes in the affirmative, otherwise a new round begins.

Fréchette et al. (2005a) showed that the results of laboratory tests on this game resemble those of the field—and those results are generally more equitable than the equilibrium predictions for payoff-maximizing players. Proposers often make generous offers and realize less power than predicted, and observed behavior is relatively invariant to institutional conditions (Fréchette et al., 2005a,b; Fréchette et al., 2012; Montero et al., 2008; Drouvelis et al., 2010). Observed equity seemingly resembles observations in ultimatum bargaining, but it cannot be explained by FS inequity aversion (Montero, 2007). Breitmoser and Tan (2013) obtain a similar result for demand bargaining games—where again experimental outcomes are equitable while FS inequity aversion predicts inequity—and show that reference dependent altruism explains the observations.

Our main hypothesis is that the reason for the equity of outcomes is the same in all three of these bargaining games—ultimatum bargaining, random-proposer majority bargaining, and demand bargaining—and that once the preference structure in these games is understood, it may allow us to predict behavior similarly well in the aforementioned dictator games of Engelmann and Strobel where FS inequity aversion fails to predict behavior. Our second hypothesis is that the utility discontinuity in reference dependent altruism helps to model this preference structure. The utility jump at the reference point implies that players are willing to take additional risks to reach it, and in this contexts it implies that voters reject inequitable proposals, gambling on the chance that they are recognized as proposers in the next round—as proposers, they will reach their reference point.

We distinguish two games, and in each of them, three players, $N = \{1,2,3\}$, have to divide €24 by majority decision. The smallest currency unit is .01 Euro. Using $C = 24$, the set of feasible allocations between the three players is

$$\mathbf{X} = \left\{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{x} \geq 0,\ \textstyle\sum_{i \in N} x_i \leq C,\ \forall i \in N : 100x_i \in \mathbb{N}_0\right\}. \tag{6}$$

The first game that we implement is the random-proposer game with a continuation probability of .95 after each round without agreement.

Game 1 (*PB95*)**.** In each round, one player is recognized as proposer by a uniform draw from N. This player chooses $\mathbf{x} \in \mathbf{X}$, and the other players vote on $\mathbf{x}$. If one of them accepts, then the players' payoffs are $\mathbf{x}$. Otherwise, the payoffs are $\mathbf{0}$ with probability .05 and a new round begins with probability .95.

PB95 is outcome equivalent to the random-proposer game with infinite time horizon and discount factor $\delta = 0.95$ if the players are risk neutral. This game has a plethora of subgame-perfect equilibria (Baron and Ferejohn, 1989), akin to folk theorems in repeated games, but analyses generally focus on equilibria in stationary strategies. Stationary strategies are independent of proposals and votes in previous rounds, and as such they are the least complex equilibrium strategies (Baron and Kalai, 1993) and imply uniqueness of ex-ante equilibrium payoffs (Eraslan, 2002).[5] Ex-ante, prior to proposer recognition, every player expects a payoff of $C/3 = 8$ under stationary

[5] Building on the assumption of stationarity, the random-proposer model has been extended in a variety of dimensions. Examples include one-dimensional ideological decisions (Cho and Duggan,

subgame perfection. Thus, payoff-maximizing voters accept any proposal that allocates them at least their "continuation payoff" $\delta 8 = 7.60$, which in turn are the costs of buying a vote. Payoff-maximizing proposers buy one vote and allocate the rest $16.40 = 24 - 7.60$ to themselves. Along the equilibrium path, proposals thus have the structure $(16.4, 7.6, 0)$ and are accepted immediately.

The second game implemented in our experiment is a random proposer game identical to *PB95* with the difference that it ends after one round, and if the first proposal is not accepted then players are paid their continuation payoffs from *PB95*. Hence, "*PB00*" is strategically equivalent to *PB95* for payoff-maximizing players, but predictions differ if players have social preferences.

Game 2 (*PB00*)**.** A player is recognized as proposer by a uniform draw from N. This player chooses $\mathbf{x} \in \mathbf{X}$, and the other players vote on $\mathbf{x}$. If one of them accepts, then the players' payoffs are $\mathbf{x}$. Otherwise, the payoffs are 7.60 per player.

If players maximize expected payoffs, the set of SPEs of *PB00* corresponds with the set of SSPEs of *PB95* in the sense that equal proposal and voting decisions are made. The ex-post payoff profile has the structure $(16.4, 7.6, 0)$ in both games.

Predictions for *PB95* and *PB00* diverge in opposite directions if players have social preferences other than RDA, and this allows us to examine the shape of their preferences. Figure 4 illustrates the ranges of equilibrium proposals compatible with the four families of social preferences. Indeed, the predicted ranges of equilibrium proposals hardly overlap, which yields the qualitative separation of utility theories that we exploit with our analysis.

To understand the divergence of predictions, let $U : \mathbb{R}^3_+ \to \mathbb{R}$ denote the utility function and assume that instead of being $U_i(\mathbf{x}) = x_i$, it is the FS utility function defined in Eq. (4). If guilt is limited as usual, $\beta < 1/2$, all proposers pay the value y that is necessary to buy one vote and keep the rest to themselves. As a result, equilibrium proposals have the structure $(24 - y, y, 0)$, where y is the transfer necessary to buy a vote. In equilibrium, the utility of the recipient of this transfer equates with his continuation utility, and assuming stationarity, and the equilibrium transfer y in *PB95*

2003; Cardona and Ponsatí, 2007), decisions with both ideological and distributive dimensions (Jackson and Moselle, 2002), bicameral legislatures (Ansolabehere et al., 2003), weighted voting (Snyder Jr et al., 2005; Montero, 2006), and costly recognition (Yildirim, 2007).

is therefore characterized by this simple condition.[6]

$$\textit{PB95:}\quad U(y,C-y,0)=\frac{\delta}{3}\big(U(C-y,y,0)+U(y,C-y,0)+U(0,C-y,y)\big)\\ +(1-\delta)\,U(0,0,0)\quad (7)$$

In contrast, the equilibrium transfer in *PB00* satisfies the condition

$$\textit{PB00:}\quad U(y,C-y,0)=U(\delta C/3,\delta C/3,\delta C/3).\qquad (8)$$

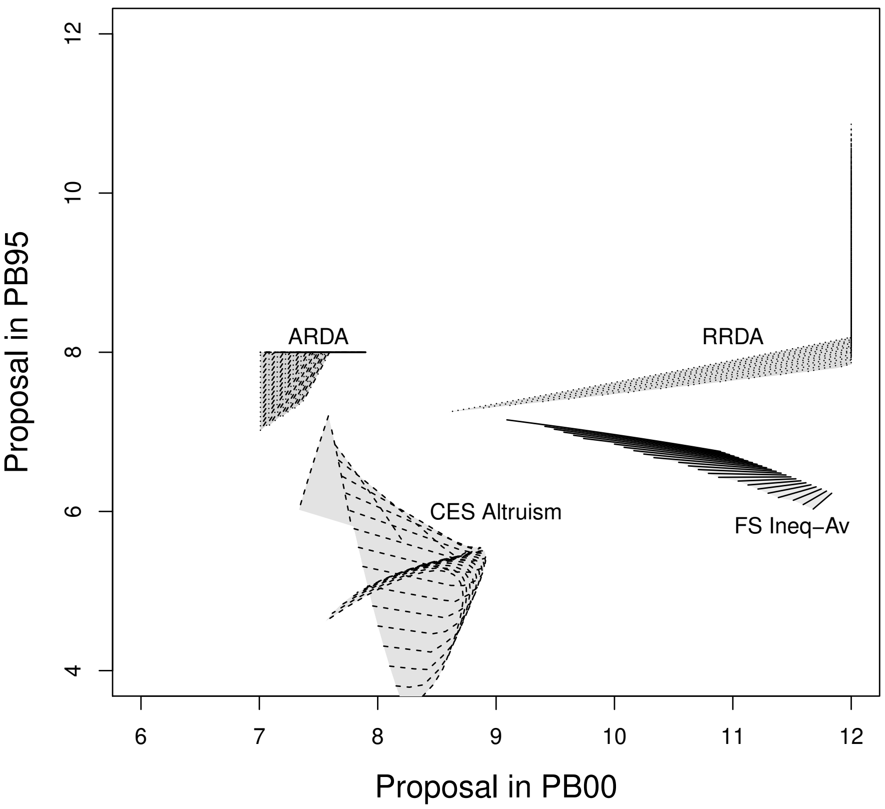
These conditions have the same solution if players maximize payoffs, $U_i(\mathbf{x}) = x_i$. If the utility is piecewise linear (as in FS inequity aversion) or non-linear (CES altruism), then the continuation utility in *PB95* (the right-hand side of Eq. (7)) differs from $U(\delta C/3,\delta C/3,\delta C/3)$ even for $y = \delta C/3$. Due to the weak concavity of U^{FS}, it is less than $U(\delta C/3,\delta C/3,\delta C/3)$, and hence, the costs of vote buying are smaller in *PB95* than in *PB00*. Solving for y under FS inequity aversion yields

$$y_{95}=C\cdot\frac{3\alpha+\delta(1-2\beta-2\alpha)}{3+6\alpha-3\beta-2\delta\cdot(\alpha+\beta)}\qquad\qquad y_{00}=C\cdot\frac{\delta/3+\alpha}{1+2\alpha-\beta},$$

and thus $y_{95} < 7.60 = \delta C/3$ and $y_{00} > 7.60 = \delta C/3$ if $0 < \beta < \alpha$ with $\beta < 1/2$. That is, inequity averse players make less equitable transfers than payoff maximizers in *PB95* and more equitable transfers than payoff maximizers in *PB00* (the former has first been observed by Montero, 2007). This applies similarly for all utility functions U that are weakly concave in the payoff profile, such as CES preferences, but the predictions of FS inequity aversion and CES altruism differ quantitatively, as shown in Figure 4. The equilibrium transfer in *PB95* drops stronger if CES altruism increases than if FS inequity aversion increases, while the transfer increases stronger in *PB00* for FS inequity aversion than for CES altruism. Assuming that proposers waste no part of the surplus, they do not "misbehave" as defined in CR reciprocity. Thus, negative reciprocity is irrelevant, rendering the specifications of FS inequity aversion and CR reciprocity equivalent. FS inequity aversion and CR reciprocity thus make the *equiva-*

[6] A standard continuity argument implies that at least one voter accepts in case of indifference. Assume there is an equilibrium where both voters reject in case of indifference. Then the proposer must offer y such that $U(y,C-y,0) > \tilde{u}$, but there is no optimal y in this case, hence no equilibrium. In turn, it is clear that there is an equilibrium where the voters accept in case of indifference, as unilateral deviations are not profitable when one is indifferent.

Figure 4: The ranges of proposals that are compatible with the four utility theories



Note: Displayed are the predicted proposals to the player whose vote is bought (for $\alpha \geq 0.1$, as the theories are indistinguishable from egoism otherwise). Specifically, CES altruism Eq. (3) for $\alpha \in [.1,.5]$ and $\beta \in [.1,.9]$, FS inequity aversion or CR reciprocity Eq. (4) for $\alpha \in [.1,1]$ and $\beta \in [.1,.33]$, absolute reference dependence Eq. (1) for for $\alpha \in [.16,.66]$ and $\beta \in [-.33,.33]$, and relative reference dependence Eq. (2) for $\alpha \in [-.33,-.1]$ and $\beta \in [-.88,-.44]$. Note that proposal range compatible with ARDA degenerates to a point in stationary SPEs of PB95 for all $\alpha - \beta > 1/4$.

lent predictions.

Intuitively, however, one may expect similarly equitable transfers in both games, *PB00 and PB95*, and indeed equitable transfers in a random-proposer game similar to *PB95* have been observed by Fréchette et al. (2005a,b). Reference dependent altruism predicts equitable transfers in *both* games for *both* reference points. Due to the utility jump at the reference point, the "bonus", the utility function is not weakly concave and the above argument no longer applies. Since the players are above their reference points at least when they are recognized as proposers, the bonus raises their continuation utility, and hence they require larger compensation from proposals that do not meet their reference point—otherwise they reject the proposal, gambling on the chance of being recognized as proposer in the next round. If players use the “relative” reference points provided by their opponents payoffs, for example, the voters require a comparably large compensation as long as the proposer gets more than them. Solving the equilibrium conditions (7) and (8) for y yields in case of RRDA

$$y_{95} = \frac{(3\alpha + (1-2\alpha)\,\delta)\,C}{(\beta-\alpha)\,\delta + 3\,(\alpha+1)} \qquad\qquad y_{00} = \frac{(3\alpha + \delta\,(1-2\beta))\,C}{3\,(\alpha+1)}. \tag{9}$$

Both y_{00} and y_{95} are greater than $\delta C/3$ for all $\alpha > \beta$ (assuming $\beta > -1$). This prediction differs qualitatively from the predictions of weakly concave utility functions and is compatible with the observations of Fréchette et al. (2005a,b). Further, as the emotional bonus $\alpha - \beta$ of reaching the reference point increases, the vote buying costs y increase further, up to $y = 12$ in *PB00*.

Finally, if players have the absolute reference point equal to their ex-ante payoff expectation, they accept (under most parameter constellations) any proposal that allocates them at least their ex-ante expectation. If $\alpha - \beta$ is not too small, they would reject any other proposal and the equilibrium proposal is $y = 8$ in *PB95*. The result is similar in *PB00*, where the equilibrium proposal can be shown to satisfy $7.6 < y < 8$ for a wide range of parameters, e.g. for all $\delta < 1$ and $\beta < 0$. Thus, ARDA predicts proposals around $y = 8$ in both games.

Experimental logistics

The experiment was conducted in the experimental economics laboratory at the Europa Universität Viadrina, Frankfurt (Oder), Germany. The experiment was, apart from the experimental instructions and control questionnaire, fully computerized (using z-Tree, see Fischbacher, 2007). Subjects were students from various faculties of the university. An announcement for this experiment was sent to recipients on an email database of potential subjects. Those who responded to the email were recruited accordingly. We conducted a total of nine sessions, four sessions of the *PB00* and five sessions of *PB95*. Each session was partitioned into two sub-sessions, to each six subjects were randomly assigned. Subjects never interacted with those from other sub-sessions. We partitioned the sessions to increase the number of independent observations, and ran them simultaneously to increase the sense of anonymity. Each session contained 12 subjects. A total of 108 subjects participated. Each subject was allowed to participate only once.

Each session comprised 10 repetitions ("stages") of the game, each comprising a number of "rounds." In each stage, subjects were randomly re-matched into groups of three, so as to implement the one-shot context. Subjects were also randomly reassigned their roles at the beginning of each round. Repetition of tasks allows for experience, while random re-matching and anonymity eliminate reputation effects. This between-subject design reduces the potential carryover effects from playing one game to another. The subjects' tasks and information during games matched precisely with the games' definitions provided above. After each stage, all subjects were informed of their earnings in that stage. Neutral language was used throughout the experiment (e.g. "A-participant" and "B-participant" instead of proposer and responder). The instructions used in *PB95* sessions are provided as supplementary material.

At the beginning of the experiment, subjects were randomly assigned computer terminals. They started by reading the experimental instructions, provided on printed sheets, followed by answering a short control questionnaire that allowed us to check their understanding. Subjects in doubt were verbally advised by the experimental assistants before being allowed to begin. Each computer terminal was partitioned, so that subjects were unable to communicate via audio or visual signals, or to look at other computer screens. Decisions were thus made in privacy. At the end of the experiment, subjects were informed of their payments, and asked to privately choose a code-name and password. This was used to anonymously collect their payments from a third party one week after the experiment. Each subject was given a € 4 participation fee and the earnings from one randomly chosen stage. The marginal incentives could therefore range from € 0 to € 24 per subject. The average payout was above € 11 per subject for, on average, less than 1 hour per session.[7]

4 The results

In this section, we analyze the qualitative compatibility of the experimental observations with the predictions of the different utility theories. Proposals are denoted as (x_p, x_h, x_l), where x_p is the proposer's payoff, $x_h := \max\{x_1, x_2\}$ is the higher of

[7] The monetary incentives provided in our experiment are substantial by local standards. Our mean payment of above € 11 per hour is, for example, 50% more than the mean wage of a research assistant.

Table 2: Means (and standard errors) of the proposals for first and second half of experiment

	Proposer payoff x_p		Higher payoff x_h		Lower payoff x_l	
	G 1–5	G 6–10	G 1–5	G 6–10	G 1–5	G 6–10
PB95	10.266 (0.5465)	10.992 (0.6411)	8.365 (0.3369)	8.911 (0.3976)	4.676 (0.5542)	3.548 (0.5554)
PB00	9.57 (0.7255)	10.899 (0.531)	8.273 (0.5503)	9.403 (0.3458)	4.887 (0.6133)	3.484 (0.6421)

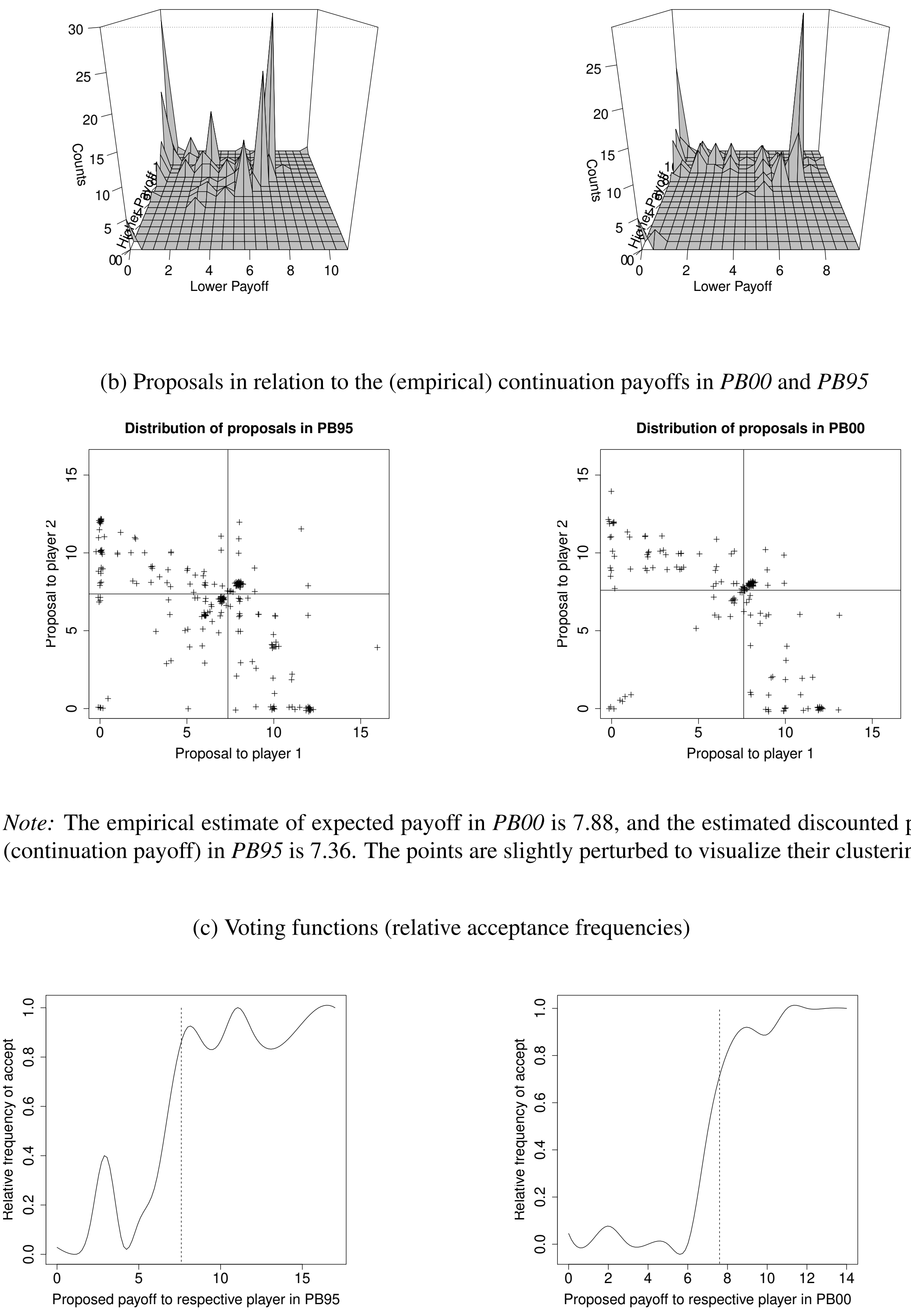
Note: The standard errors are computed using the sub-session means as independent observations. The values for "G 1–5" refer to the first five games per session, those for "G 6–10" refer to the last five games per session.

the voters' payoffs, and $x_l = \min\{x_1, x_2\}$ is the lower of the voters' payoffs. Table 2 shows that average payoffs to both co-players are higher than the $(16.4, 7.6, 0)$ predicted by egoism. The average values observed are $(10.623, 8.601, 4.155)$ for *PB95* and $(10.234, 8.838, 4.186)$ for *PB00*. Further, Mann-Whitney U tests taking the average of each session as an independent observation show that each proposal component x_p, x_h, and x_l is not significantly different across *PB95* and *PB00* ($p = 0.633$ for x_p, $p = 0.696$ for x_h, $p = 0.965$ for x_l). This holds robustly in both the first and the second half of the experiment. Recall that the non-RDA models predict $x_h < 7.6$ in *PB95* and $x_h > 7.6$ in *PB00*. The experimental observations are therefore incompatible with CES altruism, and FS inequity aversion or CR reciprocity. They are compatible with RDA, which predicts that outcomes are equitable and similarly so in *PB95* and *PB00*.[8]

Figure 5 plots the distributions of proposals in *PB95* and *PB00* from various complementary perspectives. Figure 5a plots the frequencies of proposals (x_h, x_l). Figure 5b plots the distributions of proposals made to each of the two voters. These distributions are plotted in relation to the empirical continuation payoffs, which are 7.36 in *PB95* and 7.60 in *PB00*. The plots include the proposals that were not accepted, which are located mostly in the lower-left quadrant. The vast majority of proposals is in the other three quadrants, where at least one voter's continuation payoff is met. These proposals had mostly been accepted; Figure 5c shows that offering at least one opponent a payoff of 8 ensures acceptance with high probability. We can see in Figure 5b that the proposals in both treatments are located along a perturbed concave frontier stretching

[8] Regression analyses of player-specific payoffs controlling for game (for both treatments) and round (for *PB*95) confirm the above, and also show that stationarity and truncation consistency are not violated. They are not reported here and available from the authors on request.

Figure 5: The distribution of proposals and the voting decisions



(a) Left: *PB95*, Right: *PB00*

(b) Proposals in relation to the (empirical) continuation payoffs in *PB00* and *PB95*

Note: The empirical estimate of expected payoff in *PB00* is 7.88, and the estimated discounted payoff (continuation payoff) in *PB95* is 7.36. The points are slightly perturbed to visualize their clustering.

(c) Voting functions (relative acceptance frequencies) from around $(10,0)$ through $(8,8)$ to $(0,10)$. The distributions have two mass points. In *PB95*, around 30 proposals are at $(12,0)$ or $(0,12)$, and another 30 proposals are at $(8,8)$ (see Figure 5a). In *PB00*, the mass point at $(12,0)$ is less populated. Further observations are clustered near these mass points: there is a cluster of proposals allocating 9–12 to one opponent and zero or negligible amounts to the other one and a second area to the southwest of $(8,8)$ (see Figure 5b).

ARDA is compatible with the observed treatment-invariance and average proposals, in particular of the form $(16,8,0)$, which are reflected as $(8,0)$ and $(0,8)$ in Figure 5b. Taking noise into account, ARDA is thus compatible with the observations in the cluster around $(10,0)$ in Figure 5b. We ascertain this econometrically in the next section. ARDA is incompatible with the observations around $(8,8)$, suggesting subject heterogeneity. Most observations near $(8,8)$ in Figure 5b are to its southwest, and hardly any are to its northeast. CES altruism cannot explain the cluster pattern: assuming CES altruism with $\alpha = 1/3$ ignores such asymmetry altogether, i.e. deviating from the proposal $(8,8,8)$ involves equal utility losses in either direction, be it toward $(9,8,7)$ or $(7,8,9)$. Similarly, FS inequity aversion involves continuous utility, and thus, while being asymmetric with respect to the possible deviations, it does not predict such a sharp directional effect in the neighborhood of $(8,8)$. In contrast, discontinuous utilities of RDA predict sharp effects due to jump discontinuities at reference points. With ARDA, proposers will unlikely propose $x_p < 8$ for themselves, as this would land on the wrong side of their reference point of 8 and thus yield a substantial utility drop. Hence, ARDA proposals would mostly be below the counter-diagonal through $(8,8)$, but not necessarily to its southwest. RRDA, finally, predicts that proposers unlikely propose $x_p < \max\{x_1, x_2\}$, as they are altruistic only as long as they get at least as much as either opponent, which explains the observations to the southwest of $(8,8)$.

Thus, on a qualitative basis, ARDA seems to be most compatible with the clusters around $(10,0)$ and $(0,10)$, and RRDA is most compatible with the clusters around $(8,8)$. In either case, the voters theoretically accept proposals if $y \approx 8$, which is also empirically satisfied (Figure 5c). In turn, the observations are qualitatively incompatible with CES altruism and FS inequity aversion.

5 Estimation of the utility functions

In order to verify whether RDA also fits quantitatively, we estimate the utility functions using a structural model of majority bargaining. The structural model is that of quantal response equilibrium (QRE, McKelvey and Palfrey, 1995), which relaxes the assumption of "best responses" toward "better responses" (better responses are chosen with higher probability). Specifically, we analyze *PB00*-choices through agent logit equilibrium (McKelvey and Palfrey, 1998) and *PB95*-choices through stationary logit equilibrium (Breitmoser et al., 2010), which we jointly abbreviate as SLE. Due to the large strategy sets,[9] we will also consider a generalization called stationary nested logit equilibrium (SNLE). This allows us to control for the possibility that subjects choose proposals in multiple decision steps. The clustering displayed in Figure 5b suggests that this is a possibility, and in particular it suggests that subjects first determine how many opponents and whom they pay their continuation payoffs (to buy the vote) before they choose the actual payoff profile. Hierarchical choice can be modeled using nested logit responses as defined in McFadden (1978, 1984), and in order to be on the save side, we control for this possibility. The main results do not depend on the adopted choice model, however, as RDA fits best either way. The technical details of nested logit in stationary equilibria are provided in the supplementary material.

In the previous section, we interpreted the observations to be generated by a subject pool with two discrete components (ARDA and RRDA). We model subject heterogeneity of this discrete nature using finite mixture models (McLachlan and Peel, 2000), following the literature inspired by Stahl and Wilson (1995). This allows us to simultaneously estimate number, weights, as well as utility and choice parameters of subject types. To define the likelihood function, let K denote the set of components in the population with weights ν_k and behavioral parameter tuple $\mathbf{p}^k$ for all $k \in K$. Thus with $\mathbf{P} = (\mathbf{p}^k)_{k \in K}$ as the behavioral parameter profile, and with $\mathbf{O} = (o_{s,t})$ as the set of

[9] Our analysis uses a smallest currency unit of €0.2, and given the cake sizes of €24, this implies that the number of possible proposals is on the order of 10^6 in each round. The programs underlying our computations are available as supplementary material. Analyzing such bargaining games using quantal response equilibria seems to be novel. To our knowledge, the only related analyses are Battaglini and Palfrey (2012), who study dynamic majority bargaining where the proposals are generated randomly (rather than being strategic choices), and Diermeier et al. (2002, 2003), who analyzed structural models of government formation assuming rationality during the actual bargaining phase.

observations for all subjects $s \in S$ and periods $t \in T$, the log-likelihood is

$$LL(\mathbf{P}|\mathbf{O}) = \sum_{s\in S} \ln \sum_{k\in K} \nu_k L(s,k) \qquad \text{with} \quad L(s,k) = \prod_{t\in T} \sigma\big(o_{s,t}|\mathbf{p}^k\big), \tag{10}$$

using $\sigma\left(o_{s,t}|\mathbf{p}^k\right)$ as the probability of action $o_{s,t}$ according to the QRE defined by the parameter profile $\mathbf{p}^k$. The log-likelihood is maximized jointly over all parameters to obtain consistent and efficient estimates (see e.g. Amemiya, 1978, and Arcidiacono and Jones, 2003, for further discussion), and to allow us to extract standard errors from the information matrix.[10] Model evaluation will be based on nested/non-nested likelihood-ratio tests following Vuong (1989) and on entropy-based information criteria. Such criteria extend Bayes Information Criterion $BIC = -LL + d/2 \cdot \log(O)$ (Schwarz, 1978, with number of parameters d and number of observations O) by including the entropy of posterior component membership to penalize mixture models with superfluous components. This resolves the issue that BIC overestimates the number of components of finite mixture models (Biernacki et al., 1999, 2000). The entropy-based criterion used in our analysis is the integrated classification likelihood

$$\begin{aligned} ICL\text{-}BIC &= -LL + d/2 \cdot \ln O + \mathrm{En}(\hat{\tau}) \\ &\qquad \text{with } \mathrm{En}(\hat{\tau}) = -\sum_{s\in S}\sum_{k\in K} \hat{\tau}_{sk} \ln \hat{\tau}_{sk} \quad \text{with} \quad \hat{\tau}_{sk} = \frac{\nu_k L(s,k)}{\sum_{k'\in K} \nu_{k'} L(s,k')}. \end{aligned} \tag{11}$$

Results

Table 3 presents the results. The main result, the parameter estimates of the best-fitting models are provided in Table 3a. We find that the subject pool consists of at least two components, where 55% of the subjects have RRDA preferences and 45% of them have ARDA preferences. The rather balanced distribution of RRDA and ARDA types corresponds with the previous observations that the two cluster areas contain similar numbers of observations. The RRDA component can be further split down into a sub-component with RRDA preferences (48.1%) and a sub-component with FS inequity aversion containing 7.2% of the subjects. This confirms the above qualitative observa-

[10] We use the derivative-free NEWUOA algorithm (Powell, 2008) for the initial approach toward the maximum (NEWUOA is a comparably efficient and robust algorithm, see Auger et al., 2009, and Moré and Wild, 2009), and subsequently, we use a Newton-Raphson algorithm to ensure local convergence. This procedure has been repeated using a variety of starting values. The complete list of parameter estimates is provided as supplementary material.

tion that the observations are compatible with reference dependent altruism and largely incompatible with CES altruism or FS inequity aversion. The estimated RRDA and ARDA parameters also correspond with the observations made in the previous section. The RRDA component has $\beta \approx -1$, which means that RRDA players are approximately welfare-concerned as long as they get at least as much as their opponents. This yields the cluster around $(8,8,8)$. The ARDA player have a large difference $\alpha - \beta$, which yields the treatment invariance as discussed above.

The remaining tables contain the results of our robustness checks. All parameter estimates are provided as supplementary material. We perform three sets of robustness tests, as summarized in the Tables 3b, 3c, and 3d, respectively. First, we verify whether the main results would change if we use logit instead of nested logit. Table 3b shows that the best-fitting stationary logit equilibria (SLEs) are based on ARDA and RRDA, as in the SNLE analysis, while the goodness-of-fit drops by about 500 points on the log-likelihood scale (for all utility models). Thus, acknowledging the possibility of hierarchical choice as in SNLE does not change qualitative results on the identified social motives, but it improves fit enormously. Considering the size of the strategy set, the clustering of observations, and the fact that options are simply not multinomial in majority bargaining, the tendency was expected, but the quantitative improvement is surprisingly large. It implies that the nested logit model where subjects first pick whom to pay the continuation payoff, which intuitively fits choice in majority bargaining, is much more likely to yield reliable (i.e. unbiased) estimates of the utility parameters in our context. Table 3b also informs on the goodness-of-fit of models assuming all components are either CES, FS, ARDA, or RRDA. These models fit substantially worse (at least 50 points by *ICL-BIC*) than the RRDA-ARDA mixture in Table 3a, and that FS inequity aversion and CES altruism fit substantially worse even with three components (at least 250 points by *ICL-BIC*).

Secondly, we check whether the RRDA-ARDA mixture indeed fits better than all other two-component models, even when we allow for mixtures of motives. Table 3c reports the results. It shows that regardless how the first component is modeled, a second component with FS inequity aversion fits better than CES (in terms of *ICL-BIC*), RRDA fits significantly better than FS inequity aversion ($p < .01$ in all cases), and ARDA and RRDA fit about similarly as second component. Thus, at least one component would have to be either RRDA or ARDA. Using either of ARDA and RRDA as first component, we find that one complements the other best, as reported above,

Table 3: The estimation results

(a) Estimates for the two best-fitting models

Component	Weight	λ_p	ρ_1	ρ_2	λ_v	α	β	$ICL/LL/R^2$
RRDA	0.552 (−)	5.542 (0.093)	0.145 (0.004)	0.101 (0.001)	0.001 (0.002)	−0.27 (0.002)	−0.998 (0.007)	3415.11 −3346.74 0.8914
ARDA	0.448 (0.054)	3.332 (0.042)	0.119 (0)	0.088 (0)	0.317 (0.002)	0.795 (0.006)	0.334 (0)	
RRDA	0.481 (−)	5.653 (0.027)	0.127 (0.001)	0.094 (0)	0 (0)	−0.277 (0.002)	−0.996 (0)	3405.07 −3306.89 0.8958
ARDA	0.447 (0.055)	4.604 (0.029)	0.148 (0)	0.024 (0.001)	0.328 (0.002)	0.732 (0.003)	0.36 (0)	
IneqAv	0.072 (0.023)	0.894 (0.011)	0.157 (0.006)	0.011 (0.003)	0.498 (0.01)	0.004 (0)	0.056 (0.001)	

Note: (α, β) are the parameters of the four utility functions, the remaining parameters are the choice parameters discussed in the Appendix. The standard errors are provided in parentheses. The Cox-Snell Pseudo-R^2 is $R^2 = 1 - (L(M_{\text{Baseline}})/L(M_{\text{Full}}))^{2/O}$, with the "baseline model" being the benchmark that players randomize uniformly in all cases and O being the number of observations.

(b) Goodness-of-fit (*ICL-BIC*) of logit (SLE) vs. nested logit (SNLE)

	Number of components						
Utility function	SLE × 1		SNLE × 1		SNLE × 2		SNLE × 3
CES Altr	4513.13	$\lll$	3992.96	$\lll$	3890.92	$\lll$	3702.54
IneqAv	4570.67	$\lll$	3835.85	$\lll$	3700.56	$\lll$	3666.14
RRDA	4131.97	$\lll$	3668.32	$\lll$	3469.22	=	3486.81
ARDA	4344.01	$\lll$	3621.68	$\lll$	3488.86	=	3512.42

Note: CES altruism Eq. (3), inequity aversion Eq. (4), absolute reference dependence Eq. (1), and relative reference dependence Eq. (2). The parameter estimates are supplementary material.

(c) Goodness-of-fit (*ICL-BIC*) of mixture models with two differing motives

	Second component						
First component	CES Altr		IneqAv		RRDA		ARDA
CES Altr	3890.93	$\ll$	3730.59	$\ll$	3591.86	=	3607.17
IneqAv	3730.81	=	3701.2	$\lll$	3534.5	=	3524.85
RRDA	3591.48	=	3534.87	$\ll$	3469.22	$<$	3415.11
ARDA	3607.18	$<$	3524.8	$\ll$	3415.11	$>$	3488.85

(d) Goodness-of-fit (*ICL-BIC*) of mixture models with three different components

	Third component						
First two components	CES Altr		IneqAv		RRDA		ARDA
CES + IneqAv	3744.59	=	3750.76	$\lll$	3513.81	=	3533.84
CES + RRDA	3500.23	=	3513.52	=	3531.01	$\lll$	3423.75
Ineq + ARDA	3533.81	=	3517.12	$\lll$	3404.32	$\ggg$	3537
RRDA + ARDA	3423.74	=	3405.07	=	3422.73	=	3437.96

Note: Tables (b)–(d) display the *ICL-BIC* criteria of model fit, Eq. (11), and the results of nested/non-nested Vuong tests on *ICL-BIC* for adjacent models (following the suggestion of Vuong, 1989, Eq. 5.9, we perform likelihood ratio tests including the BIC correction term and the model entropy $\text{En}(\hat{\tau})$). The signs "$<, \ll, \lll$" indicate significant improvements at $\alpha = .1, .01, .001$, respectively (note that "less is better" if goodness-of-fit is measured by information criteria such as *ICL-BIC*).

and the differences to the alternative combinations are highly substantial in terms of *ICL-BIC* (at least 100 points on the log-likelihood scale). Finally, we verify whether a possible third component (however small) is indeed best modeled by FS inequity aversion. In total, we estimate 16 three-component models, and Table 3d reports the results. All mixtures not including both RRDA and ARDA components have *ICL-BIC* criteria above 3500 points, which confirms the above results. The best-fitting three-component model, and indeed the only model that improves upon the pure RRDA + ARDA mixture in terms of *ICL-BIC*, identifies a third component of subjects with FS inequity aversion. The parameter estimates (Table 3a) show that the RRDA component is split up into two components, into one of RRDA and one of FS inequity aversion. The share of FS subjects is significant in relation to its standard error and in Vuong likelihood-ratio tests ($p < .01$), but overall it is small (7.2%). Thus, we confirm the qualitive result that reference dependent altruism fits majority bargaining behavior.

6 RDA tested out-of-sample

In this section, the validity of reference dependent altruism as a theory of social preference is tested by evaluating its predictions in a wide domain of games out-of-sample. To this purpose, we refer to experimental games of Charness and Rabin (2002, CR02), Engelmann and Strobel (2004, ES04), and Bolton and Ockenfels (2006, BO06) specifically designed to test social preferences. RDA is compared with self interest and two other social preference theories which have previously been shown to explain behavioral patterns across many games, namely FS inequity aversion, Eq. (4), and CR reciprocity, Eq. (5).

The "simple distribution experiments" of Engelmann and Strobel (2004) consist of 11 three-person dictator games of three types: taxation games, envy games, and rich-poor games. Taxation games were designed to compare the relevance of two theories of inequity aversion, namely ERC (Bolton and Ockenfels, 2000) and FS, while allowing efficiency concerns and maximin preferences as modeled by Charness and Rabin (2002). Dictators choose between three allocations, one which is predicted by ERC and another by FS inequity aversion—in half of the games efficiency or maximin predicts the same as ERC and in the other half efficiency or maximin predicts the same as FS inequity aversion. "Envy games" further test the robustness of efficiency concerns by having dictators choose between inequitable but efficient allocations versus equitable but inefficient allocations, as do "rich-poor games" which additionally are neutral to maximin preferences.

In the three-person "voting games" of Bolton and Ockenfels (2006), allocations are determined by majority vote. There are two treatments: in the "straight mode", subjects knew their roles prior to voting, and in the "equal opportunity mode", one's actual role was unknown prior to voting and there was an equal chance of being allocated to each role. Each player chooses between an equitable allocation $(13,13,13)$ versus an efficient allocation $(19,13,13)$ in Game I, $(27,1,17)$ in Game II, or $(27,9,9)$ in Game III. Relative to individual payoffs under the equitable allocation, the efficient allocation entails personal losses to none, majority, and minority of the players in Games I, II and III, respectively. Personal losses are larger in Game II than in Game III. Voting games test if there is a tradeoff between equity and efficiency.

The "simple tests" of Charness and Rabin (2002) consist of 32 games: dictator games with two or three persons, and sequential-move response games with two or three persons. In response games, the first mover chooses whether to stop the game or to let the second mover choose. The second mover's payoffs are identical across choices in some games, and in others the second mover's sacrifice helps or hurts the first mover.[11] Response games allow for tests of reciprocity, in addition to tests of distributional and welfare concerns allowed by dictator games.

With each model, we make predictions for each of the games and roles using available parameter estimates. Here, we refer to models assuming subject heterogeneity as "heterogeneous models" and to models assuming homogeneous subject pools as "homogeneous models". Besides RDA, which is a heterogeneous model, we also report predictions based on ARDA or RRDA separately of each other, which thus are homogeneous models.[12] In addition to predictions based on egoism ("Ego") and FS inequity aversion ("IneqAv"), we also test a heterogeneous model that considers both types of subjects. This follows Fehr and Schmidt (2010), who postulate that the subject pool consists of 60% egoists and 40% inequity averse types, which have

[11] There were two games where the dictator's payoffs were unknown, and so are not analyzed here.

[12] These predictions are invariant to the set of parameter estimates chosen from Table 3a, which are estimated either with or without an additional component of inequity aversion. Reference points for ARDA are, consistent with the definition given after Eq. (1) and with the random role allocation feature of both experimental designs, based on the ex-ante expectations prior to random role allocation, i.e. the equilibrium payoff of payoff-maximizing players averaged across roles.

Table 4: Predictions for the Engelmann-Strobel, Bolton-Ockenfels and Charness-Rabin games

	Observations				Predictions (Probability of a_1)								
	#Subj	a_1	a_2	a_3	Ego	ARDA	RRDA	RecChar	CR-Full	IneqAv	RDA	FS-Full	RDA-Ego
Predictions for Dictator Games in Engelmann and Strobel (2004)													
Tax-F	68	0.84	0.1	0.06	1	1	1	1	1	1	1	1	1
Tax-E	68	0.4	0.24	0.37	0	1	0	0	0	0	0.45	0	0
Tax-Fx	30	0.87	0.07	0.07	1	1	1	1	1	1	1	1	1
Tax-Ex	30	0.4	0.17	0.43	0	1	1	0	0	0	1	0	0.4
Envy-N	30	0.7	0.27	0.03	0	1	1	1	1	0	1	0	0.4
Envy-Nx	30	0.83	0.13	0.03	1	1	1	1	1	0	1	0.6	1
Envy-Ny	30	0.77	0.13	0.1	0	1	1	0	0	0	1	0	0.4
Envy-Nyi	30	0.6	0.17	0.23	0	1	1	1	1	0	1	0	0.4
RPG-R	30	0.27	0.2	0.53	1	1	1	1	1	1	1	1	1
RPG-P	30	0.6	0.07	0.33	0	1	1	1	1	0	1	0	0.4
RPG-Ey	30	0.4	0.23	0.37	0	1	1	0	0	0	1	0	0.4
Predictions for Voting Games in Bolton and Ockenfels (2006)													
Player 1													
Straight Game I	24	0.25	0.75		0	0	0	0	0	0	0	0	0
Straight Game II	24	0.33	0.67		0	0	0	0	0	0	0	0	0
Straight Game III	24	0.21	0.79		0	0	0	0	0	0	0	0	0
Equal Game I	24	0.12	0.88		0	0	0	0	0	0	0	0	0
Equal Game II	24	0.25	0.75		0	0	0	0	0	0	0	0	0
Equal Game III	24	0.17	0.83		0	0	0	0	0	0	0	0	0
Player 2													
Straight Game II	24	0.88	0.12		1	0	1	1	1	1	0.55	1	1
Equal Game II	24	0.92	0.08		1	1	1	1	1	1	1	1	1
Player 3													
Straight Game II	24	0.38	0.62		0	0	1	0	0	1	0.55	0.4	0.4
Equal Game II	24	0.25	0.75		0	0	1	0	0	1	0.55	0.4	0.4
Players 2 and 3													
Straight Game I	24	0.48	0.52		1	0	1	1	0	1	0.55	1	1
Straight Game III	24	0.88	0.12		1	1	1	1	1	1	1	1	1
Equal Game I	24	0.17	0.83		1	0	1	1	0	1	0.55	1	1
Equal Game III	24	0.85	0.15		1	1	1	1	1	1	1	1	1
Predictions for Dictator Games in Charness and Rabin (2002)													
DG2-Berk29	26	0.31	0.69		1	0	1	0	0	1	0.55	1	1
DG2-Barc2	48	0.52	0.48		1	1	1	1	1	1	1	1	1
DG2-Berk17	32	0.5	0.5		1	1	1	1	1	1	1	1	1
DG2-Berk23	36	1	0		1	1	1	1	1	0	1	0.6	1
DG2-Barc8	36	0.67	0.33		1	0	1	1	1	1	0.55	1	1
DG2-Berk15	22	0.27	0.73		1	0	0	0	0	0	0	0.6	0.6
DG2-Berk26	32	0.78	0.22		1	1	1	1	1	0	1	0.6	1
DG3-Berk24	24	0.54	0.46		1	1	1	1	1	1	1	1	1

Note: The action labeled "a_1" corresponds with A in ES04 and BO06, and with O, L in CR; "a_2" corresponds with B in ES04 and BO06, and E, R in CR02; "a_3" corresponds with "C" in ES04. The listed predictions concern the probability of a_1; the remaining probabilities follow immediately considering that no theory uniquely predicts B in ES04. Finally, below the "Scores" the p-values of tests of the null *Score (Model)* $=$ *Score(RDA)* are provided (in two-sided, matched pairs Wilcoxon tests using the individual game scores as independent observations). Model abbreviations are defined prior to Eq. (12).

	Observations				Predictions (Probability of a_1)								
	#Subj	a_1	a_2	a_3	Ego	ARDA	RRDA	RecChar	CR-Full	IneqAv	RDA	FS-Full	RDA-Ego
Predictions for Player 1 of Response Games in Charness and Rabin (2002)													
RG2-Barc7	36	0.47	0.53		0	0	0	0	0	0	0	0	0
RG2-Barc5	36	0.39	0.61		0	0	1	1	1	1	0.55	0.4	0.4
RG2-Berk28	32	0.5	0.5		0	1	1	0	0	0	1	0	0.4
RG2-Berk32	26	0.85	0.15		1	1	0	1	1	0	0.45	0.6	0.6
RG2s-Barc3	42	0.74	0.26		1	0	0	1	1	0	0	0.6	0.6
RG2s-Barc4	42	0.83	0.17		1	1	1	1	1	0	1	0.6	1
RG2s-Berk21	36	0.47	0.53		1	0	0	1	1	0	0	0.6	0.6
RG2s-Barc6	36	0.92	0.08		1	0	1	1	1	1	0.55	1	1
RG2s-Barc9	36	0.69	0.31		1	0	0	0	0	0	0	0.6	0.6
RG2s-Berk25	32	0.62	0.38		1	0	0	0	0	0	0	0.6	0.6
RG2s-Berk19	32	0.56	0.44		1	0	0	0	0	0	0	0.6	0.6
RG2s-Berk14	22	0.68	0.32		1	1	1	1	1	0	1	0.6	1
RG2s-Barc1	44	0.96	0.04		1	0	1	1	1	1	0.55	1	1
RG2s-Berk13	22	0.86	0.14		1	0	1	1	1	1	0.55	1	1
RG2s-Berk18	32	0	1		0	0	0	0	1	0	0	0	0
RG2h-Barc11	35	0.54	0.46		0	1	0	0	0	0	0.45	0	0
RG2h-Berk32	36	0.39	0.61		0	1	0	0	0	0	0.45	0	0
RG2h-Berk27	32	0.41	0.59		0	0	0	1	1	1	0	0.4	0
RG2h-Berk31	26	0.73	0.27		0	1	1	1	1	1	1	0.4	0.4
RG2h-Berk30	26	0.77	0.23		0	1	1	1	1	0	1	0	0.4
RG3-Berk16	15	0.93	0.07		0	1	1	1	1	1	1	0.4	0.4
RG3-Berk20	21	0.95	0.05		0	1	1	1	1	1	1	0.4	0.4
Predictions for Player 2 of Response Games in Charness and Rabin (2002)													
RG2-Barc7	36	0.06	0.94		1	0	1	0.5	0.5	1	0.55	1	1
RG2-Barc5	36	0.33	0.67		1	0	1	0	0	1	0.55	1	1
RG2-Berk28	32	0.34	0.66		0	0	0	0	0	0	0	0	0
RG2-Berk32	26	0.35	0.65		0	0	0	0	0	0	0	0	0
RG2s-Barc3	42	0.62	0.38		1	0	1	1	1	1	0.55	1	1
RG2s-Barc4	42	0.62	0.38		1	1	1	1	1	1	1	1	1
RG2s-Berk21	36	0.61	0.39		1	0	1	1	1	1	0.55	1	1
RG2s-Barc6	36	0.75	0.25		1	0	1	0	0	1	0.55	1	1
RG2s-Barc9	36	0.94	0.06		1	1	1	0	0	1	1	1	1
RG2s-Berk25	32	0.81	0.19		1	1	1	0	0	1	1	1	1
RG2s-Berk19	32	0.22	0.78		1	0	0	0	0	0	0	0.6	0.6
RG2s-Berk14	22	0.45	0.55		1	1	1	1	1	0	1	0.6	1
RG2s-Barc1	44	0.93	0.07		1	0	1	0	0	1	0.55	1	1
RG2s-Berk13	22	0.82	0.18		1	0	1	0	0	1	0.55	1	1
RG2s-Berk18	32	0.44	0.56		1	1	1	1	1	0	1	0.6	1
RG2h-Barc11	35	0.89	0.11		1	1	1	1	1	1	1	1	1
RG2h-Berk32	36	0.97	0.03		1	1	1	1	1	1	1	1	1
RG2h-Berk27	32	0.91	0.09		1	1	1	1	1	0	1	0.6	1
RG2h-Berk31	26	0.88	0.12		1	1	1	1	1	0	1	0.6	1
RG2h-Berk30	26	0.88	0.12		1	1	1	1	1	0	1	0.6	1
RG3-Berk16	15	0.8	0.2		0	0	0	0	0	0	0	0	0
RG3-Berk20	21	0.86	0.14		1	1	1	0	0	1	1	1	1

Table 5: Out-of-sample fit of social preference models models (negative Quadratic scores, i.e. more is better), with *p*-values of significance in relation to RDA

	Utility models								
	Ego	ARDA	RRDA	RecChar	CR-Full	IneqAv	RDA	FS-Full	RDA-Ego
Dictator Games ES04	−440.02 (0.034)	−317.68 (1)	−321.76 (1)	−361.78 (0.201)	−361.78 (0.201)	−488.02 (0.021)	**−286.3** (−)	−444.84 (0.021)	−306.36 (1)
Voting Games BO06	−181.44 (0.281)	−184.32 (0.343)	−216.96 (0.106)	−181.44 (0.281)	−147.84 (0.892)	−216.96 (0.106)	**−142.87** (−)	−172.61 (0.787)	−172.61 (0.787)
Dictator Games CR02	−206 (0.181)	−190.48 (0.371)	−185.76 (0.371)	−166 (0.371)	−166 (0.371)	−293.6 (0.1)	**−157.19** (−)	−197.84 (0.106)	−187.34 (0.181)
Response Games CR02, Pl. 1	−518.24 (0.698)	−672.4 (0.035)	−541.28 (0.272)	−480.4 (0.838)	−544.4 (0.505)	−652.16 (0.029)	−483.96 (−)	−380.77 (0.205)	**−354.66** (0.108)
Response Games CR02, Pl. 2	−464.3 (0.141)	−516.46 (0.052)	−428.46 (0.272)	−627.42 (0.008)	−627.42 (0.008)	−547.9 (0.077)	**−343.32** (−)	−416.14 (0.363)	−434.6 (0.183)
Overall	−1810 (0.005)	−1881.34 (0.001)	−1694.22 (0.004)	−1817.04 (0.003)	−1847.44 (0.004)	−2198.64 (0)	**−1413.65** (−)	−1612.19 (0.13)	−1455.58 (0.783)

Note: Data sets are abbreviated as above: CR02 is Charness and Rabin (2002), ES04 is Engelmann and Strobel (2004), BO06 is Bolton and Ockenfels (2006). Below the Quadratic Scores, the *p*-values of tests of the Null *Score (Model) = Score (RDA)* are provided, derived from two-sided, matched pairs Wilcoxon tests using the individual game scores as independent observations. Model abbreviations are defined prior to Eq. (12). "RDA-Ego" is a mixture of 60% Egoists and 40% RRDA, as a benchmark for the respective FS mixture.

$\alpha = 2/(n-1)$ and $\beta = 0.6/(n-1)$ in Eq. (4). We refer to this heterogeneous model as "FS-Full". In Charness and Rabin (2002, Table VI), one of the best-fitting models and its respective parameters estimated is their full reciprocity model ("CR-Full") with $\rho = .424, \sigma = .023, \theta = -.111$ in Eq. (5). Its predictions are partially in-sample and pose a rather tough challenge for RDA's out-of-sample predictions. We also test predictions for CR02's reciprocal charity "RecChar" model, which neglects inequity aversion (by setting $\sigma = 0$). CR02's estimates for RecChar are $\rho = .425, \sigma = 0, \theta = -.089$, and its predictions are identical to CR-Full's in many games.

For all utility functions and all games, we derive the unique predictions without noise and evaluate their adequacy using the quadratic scoring rule (Selten, 1998; Gneiting and Raferty, 2007).[13] With G as the set of games considered here, $A(g)$ as the action set in game $g \in G$, $n(a,g)$ as the number of subjects that chose a in g, and $p(a,g)$ as the predicted probability of a in g, the prediction scores are

$$\textit{Quadratic Score:} \quad S_Q = -\sum_{g \in G} \sum_{a \in A(g)} \sum_{b \in A(g)} n(a,g) \cdot \left(I_{a=b} - p(b,g)\right)^2. \tag{12}$$

Table 5 contains the overall scores, the scores for subsets of games, and in parentheses p-values of two-sided Wilcoxon matched pair tests of differences to RDA (if $p < .05$, then the respective model fits significantly worse than RDA). Table 4 contains the predictions of all models for all games. The main results are that the heteroge-

[13] We evaluate the predictions without noise, as noise parameters such as those estimated above or by Charness and Rabin are not transferable across experiments. In case a model's prediction is indeterminate, we refine it in the sense of the respective theories. The Ego prediction is refined toward inequity aversion in cases of indeterminacy.

neous RDA model fits significantly better than all homogeneous models, and overall also better than the heterogeneous Fehr-Schmidt model FS-Full. The latter is not quite significant, but RDA fits better than FS-Full in four of the five classes of games, which we discuss in more detail shortly. The best-fitting homogeneous model is RRDA, and in fact it is the only model that improves upon Ego overall. As a homogeneous model, ARDA does not fit well, but this is unsurprising, as ARDA captures the behavior of a minority in our experiment, and thus it is not expected to extrapolate to the population as a whole. It complements RRDA well, however, as RDA fits significantly better than RRDA overall ($p = .026$).

For the dictator games of CR02 and ES04, ARDA and RRDA fit about similarly, and both of these models fit about as well as the Charness-Rabin model CR-Full. ARDA and RRDA fit slightly better in the Engelmann-Strobel games, CR-Full fits slightly better in the Charness-Rabin games. In both cases, inequity aversion and the heterogeneous FS-Full model fit poorly, i.e. these experiments indeed identify the limits of FS inequity aversion. Interestingly, the simple switch from the continuous FS utility function to the discontinuous RRDA utilities eliminates exactly these issues, and correctly predicts conditional welfare concerns as hypothesized above. This confirms that the condition for welfare concerns depends on the relation of one's payoff to the reference point.

For BO06's voting games, RDA has the highest score. In many cases, RDA makes the same prediction as all the other models, and these predictions are qualitatively in line with behavior. This explains why the overall differences between model scores are statistically insignificant for voting games. In the remaining cases where RDA's predictions are different from those of other models, RDA's predictions are non-degenerate and the corresponding observations are heterogeneous. Thus, RDA does well in capturing the equity-efficiency tradeoff BO06 speak of, e.g. for Game I players 2 and 3 as well as Game II player 3.

Finally, we turn to CR02's response games. Table 3a shows that RRDA accurately predicts the reciprocal choices of player 2, while it does poorly in predicting the strategic reciprocity of player 1. In the latter case, egoism does fairly well, and the only homogeneous model improving upon egoism in predicting strategic reciprocity is the reciprocal charity model of Charness and Rabin (Blanco et al., 2011, discuss the difficulties in predicting strategic reciprocity in more detail). Surprisingly, the heterogeneous FS-Full model does well in predicting strategic reciprocity, which can be attributed to its inclusion of egoism. To verify this, we estimated an "RDA-Ego" mixture of 60% Egoists and 40% RRDA, i.e. a mixture that substitutes FS inequity aversion (IneqAv) with RRDA, and this model fits substantially better than FS here and overall. Again, this illustrates that the discontinuous RRDA utilities captures behavior better than the continuous FS inequity aversion.

We therefore conclude that the contingency of altruism to reference points captures the interplay between inequity aversion, welfare concerns and non-strategic reciprocity very well, even out-of-sample. In turn, it does not (out-of-sample) improve on reciprocal charity (in-sample) in predicting strategic reciprocity, but it improves upon the remaining models also in this respect.

7 Concluding discussion

This paper presents a theory of reference dependent altruism. RDA extends previous theories by organizing data across a wider range of games. We estimate that 45% of subjects use their ex-ante expected payoff as reference points and that 55% of subjects use the opponents' payoffs as reference points. In either case, the altruism weight α we estimate is high if one's payoff exceeds one's reference point and it is low otherwise. It explains majority bargaining, where behavior is incompatible with existing theories. It also fits the behavioral pattern observed across a wide variety of experimental games from key studies of social preferences better than existing theories do.

RDA combines insights from existing models rather than taking an entirely different approach. Instead of modeling the interplay of equity, efficiency, and reciprocity concerns explicitly by assigning exogeneous context-dependent weights to each concern, RDA models it implicitly and thus its novelty. The altruistic utility derived from co-player payoffs models efficiency concerns, the dependence of its weight α on relative reference points models equity concerns, and the increase in α at the reference point yields reciprocity effects. The interplay of these concerns organizes the seemingly disparate behavioral pattern observed in the literature.

Most distinctively as a theory of social preferences, reference dependent altruism weights predict a utility jump at the reference point. In the context of majority bargaining, this implies that subjects are willing to take risks to reach their reference points, whereas they are risk neutral in our piecewise linear model as long as they are securely above or below it. In addition, the model predicts efficiency concerns above the reference point and reciprocity conditional on reaching it. The better off subjects are, the more altruistic they are towards others; anticipating this, others may benefit from treating them well before they move. These patterns were observed in our experiment.

Going by our estimates of α in majority bargaining, utility increases in a co-player's payoff even if it is more than one's payoff. This correctly predicts that all player 2 in the mini ultimatum games shown in Figure 1 prefer an inequitable distribution of $(8,2)$ to an inefficient distribution of $(0,0)$. FS inequity aversion predicts $(0,0)$ using the standard weight placed on envy, which Charness and Rabin in turn estimated to be negligable. Indeed, the generally low rejection rates observed in ultimatum bargaining especially after stabilizing with experience (Cooper and Dutcher, 2011) are more compatible with the predictions of CR reciprocity and RDA than of the FS inequity aversion.[14]

RDA captures equity concerns, though not through disutility from payoff differences as modeled in CR reciprocity and FS inequity aversion. Instead, a player with RRDA always derives utility from a co-player's payoff, but more so when the player's payoff is equal to or greater than the co-player's ($\alpha \approx 0.9$) than when the player's payoff is less than the co-player's payoff ($\alpha \approx 0.2$). With the relative reference point and the utility function $u_i = x_i + \alpha x_j$, the resulting utility jump at the reference point predicts the preference for distributions that are equitable in the CR02 games in Figure 2, where the reference point is reached in exactly one of the two options. Based on CR reciprocity using in-sample parameter estimates, efficiency or reciprocity concerns should counter these effects and envy should be negligable, contradicting the observed choices. RRDA correctly predicts player 2 behavior in these games.

RDA also captures efficiency concerns. Efficiency concerns dominated equity concerns in most of ES04 games. The key to understanding these results is that across options—for all but one of the games—the dictator's payoff is always higher than one co-player's and lower than the other co-player's. This implies that across options, the degree of altruism towards each co-player is constantly high or low, respectively, as reference points are never crossed. This predicts the efficiency concerns observed by Engelmann and Strobel (2004) very well, as shown in Table 5. Returning to the games in Table 1, RDA predicts well in ES04 distribution games and BO06 voting

[14] That said, CR reciprocity and RDA can also predict rejection for a range of parameter values.

game Game I where efficiency concerns dominate, and voting games Game II and III where equity concerns dominate. RDA thus captures the tradeoff between efficiency and equity concerns, as reverberated in the dialogue between Engelmann and Strobel (2006) and Bolton and Ockenfels (2006).

The potential of RDA can be exploited with a deeper understanding of reference points. Many situations have more than one reference point, as we have analyzed. Our estimates indicate a heterogeneity of subjects who use absolute or relative reference points in such environments. Different reference points could also arise from the plurality of fairness norms, and from whether the surplus bargained over was gained by windfall or was legitimized by effort (Cherry et al., 2002; Cappelen et al., 2007). RDA does not exclude the use of other reference points, and we conjecture that the focality of reference points determines their relevance. Predictions can be sharpened by identifying factors that influence the relative importance of available reference points.

References

Amemiya, T. (1978). On a two-step estimation of a multivariate logit model. *Journal of Econometrics*, 8(1):13–21.

Andreoni, J. and Bernheim, B. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Andreoni, J. and Miller, J. (2002). "Giving according to GARP: An experimental test of the consistency of preferences for altruism". *Econometrica*, 70(2):737–753.

Andreoni, J. and Sprenger, C. (2012a). Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–56.

Andreoni, J. and Sprenger, C. (2012b). Risk preferences are not time preferences. *American Economic Review*, 102(7):3357–76.

Ansolabehere, S., Snyder Jr, J., and Ting, M. (2003). Bargaining in bicameral legislatures: When and why does malapportionment matter? *American Political Science Review*, 97(3):471–481.

Arcidiacono, P. and Jones, J. (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, 71(3):933–946.

Auger, A., Hansen, N., Perez Zerpa, J., Ros, R., and Schoenauer, M. (2009). Experimental comparisons of derivative free optimization algorithms. *Experimental Algorithms*, pages 3–15.

Baron, D. and Ferejohn, J. (1989). Bargaining in legislatures. *American Political Science Review*, 83(4):1181–1206.

Baron, D. and Kalai, E. (1993). The simplest equilibrium of a majority rule division game. *Journal of Economic Theory*, 61(2):290–301.

Battaglini, M. and Palfrey, T. (2012). The dynamics of distributive politics. *Economic theory*, 49(3):739–77.

Bicchieri, C. and Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making*, 23(2):161–178.

Biernacki, C., Celeux, G., and Govaert, G. (1999). An improvement of the nec criterion for assessing the number of clusters in a mixture model. *Pattern Recognition Letters*, 20(3):267–272.

Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(7):719–725.

Blanco, M., Engelmann, D., and Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 72(2):321–338.

Bolton, G. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.

Bolton, G. E. and Ockenfels, A. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment. *The American economic review*, 96(5):1906–1911.

Breitmoser, Y., Tan, J., and Zizzo, D. (2010). Understanding perpetual R&D races. *Economic Theory*, 44(3):445–467.

Breitmoser, Y. and Tan, J. H. (2013). Reference dependent altruism in demand bargaining. *Journal of Economic Behavior and Organization*, 92:127–140.

Camerer, C. and Thaler, R. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, pages 209–219.

Cappelen, A., Hole, A., Sørensen, E., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97(3):818–827.

Cardona, D. and Ponsatí, C. (2007). Bargaining one-dimensional social choices. *Journal of Economic Theory*, 137(1):627–651.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869.

Cherry, T., Frykblom, P., and Shogren, J. (2002). Hardnose the dictator. *American Economic Review*, 92(4):1218–1221.

Cho, S. and Duggan, J. (2003). Uniqueness of stationary equilibria in a one-dimensional model of bargaining. *Journal of Economic Theory*, 113(1):118–130.

Cooper, D. J. and Dutcher, E. G. (2011). The dynamics of responder behavior in ultimatum games: a meta-study. *Experimental Economics*, 14(4):519–546.

Cox, J., Friedman, D., and Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59(1):17–45.

Diecidue, E. and Van de Ven, J. (2008). Aspiration level, probability of success and failure, and expected utility*. *International Economic Review*, 49(2):683–700.

Diermeier, D., Eraslan, H., and Merlo, A. (2002). Coalition governments and comparative constitutional design. *European Economic Review*, 46(4-5):893–907.

Diermeier, D., Eraslan, H., and Merlo, A. (2003). A structural model of government formation. *Econometrica*, 71(1):27–70.

Drouvelis, M., Montero, M., and Sefton, M. (2010). Gaining power through enlargement: Strategic foundations and experimental evidence. *Games and Economic Behavior*, 69(2):274–292.

Engelmann, D. and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, 94(4):857–869.

Engelmann, D. and Strobel, M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Reply. *The American economic review*, 96(5):1918–1923.

Eraslan, H. (2002). Uniqueness of stationary equilibrium payoffs in the baron–ferejohn model. *Journal of Economic Theory*, 103(1):11–30.

Fehr, E. and Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

Fehr, E. and Schmidt, K. (2010). On inequity aversion: a reply to binmore and shaked. *Journal of economic behavior & organization*, 73(1):101–108.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fréchette, G., Kagel, J., and Morelli, M. (2005a). Behavioral identification in coalitional bargaining: An experimental analysis of demand bargaining and alternating offers. *Econometrica*, 73(6):1893–1937.

Fréchette, G., Kagel, J., and Morelli, M. (2005b). Nominal bargaining power, selection protocol, and discounting in legislative bargaining. *Journal of Public Economics*, 89(8):1497–1517.

Fréchette, G., Kagel, J., and Morelli, M. (2012). Pork versus public goods: an experimental study of public good provision within a legislative bargaining framework. *Economic Theory*, 49:779–800.

Gneiting, T. and Raferty, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Goeree, J., Holt, C., and Laury, S. (2002). Private costs and public benefits: unraveling the effects of altruism and noisy behavior. *Journal of Public Economics*, 83(2):255–276.

Jackson, M. and Moselle, B. (2002). Coalition and party formation in a legislative voting game. *Journal of Economic Theory*, 103(1):49–87.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Knez, M. and Camerer, C. (1995). Outside options and social comparison in three-player ultimatum game experiments. *Games and Economic Behavior*, 10(1):65–94.

Knight, B. (2005). Estimating the value of proposal power. *American Economic Review*, 95(5):1639–1652.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121(4):1133–1165.

Kőszegi, B. and Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, 97(4):1047–1073.

Kritikos, A. and Bolle, F. (2001). Distributional concerns: equity- or efficiency-oriented? *Economics Letters*, 73(3):333–338.

McFadden, D. (1978). Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial interaction theory and planning models*, pages 75–96. North Holland, Amsterdam.

McFadden, D. (1984). Econometric analysis of qualitative response models. *Handbook of econometrics*, 2:1395–1457.

McKelvey, R. and Palfrey, T. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

McKelvey, R. and Palfrey, T. (1998). Quantal response equilibria for extensive form games. *Experimental Economics*, 1(1):9–41.

McLachlan, G. and Peel, D. (2000). *Finite mixture models*, volume 299. Wiley-Interscience.

Montero, M. (2006). Noncooperative foundations of the nucleolus in majority games. *Games and Economic Behavior*, 54(2):380–397.

Montero, M. (2007). Inequity aversion may increase inequity. *The Economic Journal*, 117(519):192–204.

Montero, M., Sefton, M., and Zhang, P. (2008). Enlargement and the balance of power: an experimental study. *Social Choice and Welfare*, 30(1):69–87.

Moré, J. and Wild, S. (2009). Benchmarking derivative-free optimization algorithms. *SIAM Journal on Optimization*, 20(1):172–191.

Murnighan, J., Roth, A., and Schoumaker, F. (1988). Risk aversion in bargaining: An experimental study. *Journal of Risk and Uncertainty*, 1(1):101–124.

Neilson, W. (2006). Axiomatic reference-dependence in behavior toward others and toward risk. *Economic Theory*, 28(3):681–692.

Okada, A. (2011). Coalitional bargaining games with random proposers: Theory and application. *Games and Economic Behavior*, 73(1):227 – 235.

Powell, M. (2008). Developments of newuoa for minimization without derivatives. *IMA journal of numerical analysis*, 28(4):649.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 83(5):1281–1302.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, 1(1):43–62.

Shalev, J. (2002). Loss aversion and bargaining. *Theory and Decision*, 52(3):201–232.

Snyder Jr, J., Ting, M., and Ansolabehere, S. (2005). Legislative bargaining under weighted voting. *American Economic Review*, 95(4):981–1004.

Stahl, D. and Wilson, P. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.

Thaler, R. (1988). Anomalies: The ultimatum game. *The Journal of Economic Perspectives*, 2(4):195–206.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333.

Yildirim, H. (2007). Proposal power and majority rule in multilateral bargaining with costly recognition. *Journal of Economic Theory*, 136(1):167–196.