Soc Choice Welf (2015) 45:489–511
DOI 10.1007/s00355-014-0818-7

The impact of relative position and returns on sacrifice and reciprocity: an experimental study using individual decisions

Jordi Brandts · Enrique Fatas · Ernan Haruvy · Francisco Lagos

Received: 27 March 2014 / Accepted: 1 April 2014 / Published online: 11 July 2014

Abstract We present a comprehensive experimental design that makes it possible to characterize other-regarding preferences and their relationship to the decision maker's relative position. Participants are faced with a large number of decisions involving variations in the trade-offs between own and other's payoffs, as well as in other potentially important factors like the decision maker's relative position. We find that: (1) choices are responsive to the cost of helping and hurting others; (2) The weight a decision maker places on others' monetary payoffs depends on whether the decision maker is in an advantageous or disadvantageous relative position; and (3) We find no evidence of reciprocity of the type linked to menu-dependence. The results of a mixture-model estimation show considerable heterogeneity in subjects' motivations and confirm the absence of reciprocal motives. Pure selfish behavior is the most frequently observed behavior. Among the subjects exhibiting social preferences, social-welfare maximization is the most frequent, followed by inequality-aversion and by competitiveness.

J. Brandts (✉)
Institut d'Anàlisi Econòmica (CSIC) and Barcelona GSE, Barcelona, Spain
e-mail: jordi.brandts@iae.csic.es

E. Fatas
School of Economics, University of East Anglia, Norwich NR4 7TJ, UK
e-mail: e.fatas@uea.ac.uk

E. Haruvy
School of Management, University of Texas—Dallas, Richardson, TX, USA
e-mail: eharuvy@utdallas.edu

F. Lagos
GLOBE, Departamento de Teoría e Historia Económica, Facultad de Ciencias Económicas,
Universidad de Granada, Granada, Spain
e-mail: fmlagos@ugr.es

Keywords Experiments · Social preferences · Reciprocity · Difference aversion

JEL Classification C70 · C90 · D63 · D64

1 Introduction

Studies of social preferences in economics have demonstrated that decision makers consider the welfare of others. Much of the initial scholarly debate focused on what kind of utility specification was most appropriate for characterizing people's average behavior. The specific motives explored in the literature can be roughly classified into altruistic preferences (Andreoni et al. 2008), distributional preferences (Bolton and Ockenfels 2000; Fehr and Schmidt 1999), and reciprocal preferences (Charness and Rabin 2002; Falk and Fischbacher 2006). Altruistic preferences involve a weight on the others' payoffs, distributional preferences typically involve a penalty on the distance from some equitable reference point, and reciprocal preferences are conditional on actions by others.

Extant studies of people's pro-social behavior have looked at relatively few decisions by each individual and established models of social preferences have been informed by this kind of aggregate information. However, some of the more recent evidence has shown that individual motivations are quite heterogeneous. This is important from a positive point of view, since aggregate behavior often needs to be understood in terms of the interaction of agents with different kinds of social preferences. To better understand heterogeneity is also important from a normative standpoint. Fairness ideals can be expected to be diverse and to understand public debates on distributive justice one needs to have information about this diversity. As Almas et al. (2010) put it, "most adults believe that differences in individual achievements and efficiency considerations [...] may justify an unequal distribution of income". But, they disagree on which unequal distributions they consider to be fair. From a normative point of view, understanding the heterogeneity of preferences and views may help to get a better understanding of the sources of these disagreements. Cappelen et al. (2007) analyze how views on fairness ideals depend on individual characteristics, including talent and disposition to effort. Konow (2009) documents the existence of large variance in fairness ideals in a realistic experiment in which the diversity of moral judgments is sensible to the manipulation of information conditions in the laboratory.

For both positive and normative reasons it is, therefore, necessary to generate information that directly pertains to *individuals'* social preferences. Experiments make it possible to generate rich data sets of this kind and this can yield important advantages in some cases. Some previous studies collect such data sets. Brandts and Schram (2001) study a public good environment in which subjects have to make contribution decisions for 10 different relative prices of a private and public good. This yields a complete 'contribution function' for each subject and makes it possible to reject in a simple way the long-lived hypothesis that subjects contribute positive amounts only by mistake. In Johansson-Stenman et al. (2002) subjects are asked to choose between alternative states with different uniform income distributions. Through choices between alternative states they obtain information about participants' degree of relative risk aversion and the degree of positionality (the concern for relative standing).

Andreoni and Miller (2002) ask whether experimental subjects' concern for altruism or fairness can be expressed as a well-behaved preference ordering. They had their subjects make between 8 and 16 dictator-type allocation decisions in which they distributed tokens between themselves and another person. Each of these involved a given endowment of tokens as well as an own payoff from keeping a token and a payoff to the other from giving a token to the other. Of these situations 11 involved a token endowment, a positive value to the decision-maker of keeping each token and a positive value (for the other person) of each of the tokens being passed to him or her. These eleven situations involved five different levels of token endowments and seven different values for the relative price of giving. The other five situations effectively involved budget constraints that sloped up—people could spend tokens on taking tokens away from others—with the aim of finding out about the importance of what the authors call "rational jealousy". This design makes it possible to obtain a broad picture of behavior and to study the consistency of behavior in their context. Andreoni and Miller (2002) report that a large part of their subjects' decisions can be represented by three kinds of simple distributional preferences: purely selfish, Leontief and perfect substitutes preferences.[1]

Blanco et al. (2011) study the behavior of subjects that play the ultimatum game, the dictator game, a sequential prisoner's dilemma and a public good game. Their data show that inequality aversion has good predictive power at the aggregate level but performs less well at the individual level. Iriberri and Rey-Biel (2013) ask for actions over sixteen modified dictator games with three possible choices each and elicit beliefs about other subjects' actions in the same task; their focus is on the correlation between actions and beliefs and the impact of social information. They report that subjects with different interdependent preferences have indeed different beliefs about others' actions.

Cappelen et al. (2011) use a multi-stage dictator game where the donation is preceded by a production stage in which dictators had to invest their endowment to obtain a (high or low) return. They obtain information about subjects' social preferences from a standard experiment with real stakes and self-reported information on fairness ideals coming from surveys. Relative to a base experimental condition (the B-treatment), they prime different fairness rules in their E-treatment and find a persistent heterogeneity of subjects' ideals across information conditions and experimental methods (with real stakes or using self-reported data).

In this work, we attempt to measure the impact both qualitative and quantitative of these three basic factors. First, through a series of different binary choices we elicit whether subjects are willing—and to what extent—to increase or decrease others' payoffs at different costs to their own payoffs. Second, we study how such decisions depend on whether the decision maker is in an advantageous or disadvantageous position. Third, we study whether these same subjects modify their willingness to sacrifice depending on whether another subject has foregone various types of outside

[1] Fisman et al. (2007) go further and allow for three-person dictator games. This makes it possible to distinguish between behavior that is compatible with well-behaved preferences and behavior that is not.

options (reciprocity). In this way, we present a joint analysis of the importance of distributional aspects linked only to payoffs and of reciprocity influences of the kind related to menu-dependence.[2]

In our experiments, each of our participants has to make numerous sequential pairwise choices between two alternative states described only by payoffs to himself and to another person. In designing these choice problems we were guided by what previous experimental and modeling studies have identified as some of the basic influences that need to be taken into account when dealing with interdependent motivation. The first of these is that many people care about the payoffs of others in several respects. People may care about the relation of their own payoffs to those of others, about the sum of payoffs to all involved, as well as about other payoff-related aspects. The second lesson that has been learned from previous work is that people may also be influenced by a variety of circumstances surrounding the act of choice that are not directly payoff-relevant. These circumstances may include aspects like the features of foregone payoff distributions, the procedure by which an outcome is reached and the beliefs that people hold about the intentions behind others' choices. Our design allows us to study the relevance of these different forces at the individual level.

We present a regression analysis of behavior based on the individual data. Our results show that reciprocity linked to menu-dependence is never significant in our data. We also find that subjects are influenced by the price of sacrificing money and by whether they choose from a strong or from a weak position.

We then use our data to study individual behavior more formally. Previous research has identified several possible specifications of social preferences models.[3] These different specifications can be nicely embedded in the general model presented in Charness and Rabin (2002). In particular, the model includes a parameter that captures reactions to positive and negative actions. On the basis of this model we use our data for a mixture-of-types model estimation. Our results exhibit considerable heterogeneity of types. We find a large fraction of selfish subjects. We also find that social-welfare maximizers are more frequent than inequity-averse subjects. None of the preference types that emerge involve a significant value of the parameter related to positive or negative actions.

2 Experimental setting

There are two subject roles denoted 1 and 2. Each player 1, the dictator, makes sixty six binary dictator choices between A and B.

The sixty six dictator choices can be classified into six regimes based on relative position and the recipient's previous choice. In the first two regimes, the dictator chooses from two different relative positions, *strong* and *weak*, without the intervention of the other player. In the strong position, the dictator's payoff is at least as large as

[2] Note that in our experiment all payoffs are deterministic and luck plays no role. Becker (2013) is a recent example of how different types of luck change distributional preferences.

[3] See Cooper and Kagel (2011) for a recent survey of the experimental literature on other-regarding preferences.

Table 1 Choice tasks involving a dictator with a strong position

Choice task	Choice A	Choice B	Return on sacrificing	Critical ρ if indifference between A and B when $q = 0$
1	(1100, 600) vs.	(1000,1000)	4	0.20
2	(1100, 600) vs.	(1000, 900)	3	0.25
3	(1100, 600) vs.	(1000, 800)	2	0.33
4	(1100, 600) vs.	(1000, 700)	1	0.50
5	(1100, 600) vs.	(1000, 600)	0	1
6	(1100, 600) vs.	(1000, 500)	−1	–
7	(1100, 600) vs.	(1000, 400)	−2	−1
8	(1100, 600) vs.	(1000, 300)	−3	−0.50
9	(1100, 600) vs.	(1000, 200)	−4	−0.33
10	(1100, 600) vs.	(1000, 100)	−5	−0.25
11	(1100, 600) vs.	(1000, 0)	−6	−0.20

the passive player's payoff. In contrast, in the weak position, *the dictator*'s payoffs never exceed the recipient's payoffs. We designed these two environments, motivated by the previous evidence that pointed to the relevance of relative position.

Table 1 shows the specific alternatives from a strong position. Subjects in the player 1 role had to make eleven choices between A and each of the B-choices. Each binary choice between A and B involves a return on sacrificing, shown in the fourth column of the table.[4] This variation in return across choices allows us to determine the extent to which people sacrifice their own material payoffs to increase or to decrease other people's payoffs. Consider an example. The binary choice between state A and Bin choice task 1 consists of player 1 deciding whether to forego 100 units to raise the other subject's payoff by 400 units, so that for player 1 the price of each payoff unit given to player 2 is .25. The number in the fourth column of Table 1 will be used in our figures to refer to the different Bs.

The selection of payoffs in Table 1 responds to the following considerations. Since we want to place player 1 in a strong position, player 1's payoff can be no smaller than that of player 2 in any of the possible outcomes. The relation between payoffs at A and those at any of the B cases has to be such that player 1 gives up a part of his payoff and alters that of player 2; player 1 pays a price for changing player 2's payoff. We also wanted to give player 1 both the possibility of increasing and decreasing the other player's payoff. In this we were guided by the already abundant evidence which shows that many people are willing to act in this manner.[5] All these considerations impose that player 1's payoff in all the B-choices has to be smaller than 1100 and that

[4] The returns on sacrificing are given by the player 2's gain from switching from option A to a particular option B divided by the player 1's loss from choosing any B (always 100).

[5] Zizzo and Oswald (2001) report results from an experiment in which they vary the price of burning money, i.e. the amount of their own money that subjects must give up to decrease other people's money holdings. They found that nearly two-thirds of subjects paid for impoverishing other people. Even as the price of burning went up, the percentage of people who chose to burn other people did not fall substantially.

Table 2 Choice tasks involving a dictator with a weak position

Choice task	Choice A	Choice B	Return on sacrificing	Critical σ if indifference between A and B when q = 0
12	(600, 1100) vs.	(500, 1500)	4	0.20
13	(600, 1100) vs.	(500, 1400)	3	0.25
14	(600, 1100) vs.	(500, 1300)	2	0.33
15	(600, 1100) vs.	(500, 1200)	1	0.50
16	(600, 1100) vs.	(500, 1100)	0	1
17	(600, 1100) vs.	(500, 1000)	−1	–
18	(600, 1100) vs.	(500, 900)	−2	−1
19	(600, 1100) vs.	(500, 800)	−3	−0.50
20	(600, 1100) vs.	(500, 700)	−4	−0.33
21	(600, 1100) vs.	(500, 600)	−5	−0.25
22	(600, 1100) vs.	(500, 500)	−6	−0.20

player 2's payoff in these states has to vary in a way that implies different positive and negative returns on sacrificing. Here is where we introduce a simplifying element into the design by keeping player 1's payoff always at a value of 1000 in the different B choices and changing only player 2's payoff. Given this choice, player 2's payoff in the situation most favorable to player 2, B_1, can not be higher than 1000, since otherwise player 1 would cease to be in the strong position. From here the other states are derived by diminishing player 2's payoff until we reach zero. Some of the positive and negative returns have the same absolute value. This is not a necessary feature of the design, but introduces some additional simplicity. As a final feature, note an additional connection between Tables 1 and 2: total payoffs are the same for any particular row of the two tables.

Table 2 shows the set of alternatives from a weak position for players in the *dictator* role. Note first that here the payoffs in state A are just reversed with respect to what they were for the case where choices are from a strong position. As before, when a subject chooses state B over A in any of the first five binary choices, she is sacrificing own material payoffs to *help* the other subject and when a subject chooses state B over A in any of the last six binary choices, she is sacrificing own material payoffs to *hurt* the other subject. The different B payoffs are chosen in such a way that player 1's payoff loss is the same for all the B states and that the returns on sacrificing shown in column 4 of Table 2 are the same as for the choices from a strong position. As in Table 1, fourth column shows a number for each return which will be used as a label in the graphical representations below.

The choices presented in Tables 1 and 2 are the baseline for the four remaining environments we confront our subjects with. In these four so-called response games player 2 first decides whether to accept an outside option or to let player 1 make a set of binary choices as above. This kind of response games are used extensively in Charness and Rabin (2002). Table 3 gives an overview of the four response games we used. The names of the different games are meant to capture player 1's payoff in relation to player 2. The letter “S” stands for player 1’s strong position and the letter “W” for her weak position, as used above. The labels “PR” and “NR” refer to positive and negative reciprocity. We use these terms here in a descriptive way[6] to refer to the fact that for “PR” (“NR”) player 1 obtains less (more) at the outside option than at any of the choices between A and B, and that player 1 may react favorably (unfavorably) to this fact. The responsibility associated with others’ choices can influence people’s rankings over narrowly defined outcomes and, in our context, this pertains to the available outside option.

If in the SPR response game player 2 gives up his outside option she allows player 1 to obtain either 1000 or 1100, in both cases substantially more than the 0 payoff that he would have obtained at the outside option. Observe that at the outside option player 2 obtains a payoff of 1000, so that by passing up that outside option player 2 exposes herself considerably, since 1000 is the most player 2 can get from player 1’s choices. The fact that player 2 has nothing to gain in terms of own payoff from foregoing the outside option, makes this an environment favorable to the emergence of positive reciprocity. Specifically, if player 2 allows player 1 to effectively play, then player 1 can be expected to be more generous towards player 2 than in the absence of the (0, 1000) outside option. For the SNR game one can make a similar argument; by not taking the outside option, player 2 imposes a large loss on player 1 while player 2 can still obtain the same—or a similar—payoff than at the outside option. As a consequence, if player 1 is called to play he can be expected to be less generous than in the absence of the (2000, 1000) outside option.

For the two response games involving the weak position one can say something rather analogous. Player 1’s payoff at the outside option of game WPR (WNR) is with 0 (2000) lower (higher) than any of the two payoffs that he can obtain if player 2 foregoes the outside option. Player 2’s payoff is equal to the highest possible payoff that can arise if she gives player 1 the opportunity to choose.[7]

In summary, our data consist of individual choices for what can be seen as 22 different budget set segments, involving both positive and negative relative prices, for three cases (the distributional treatment and the two variations of the reciprocity treatment) which differ with respect to the overall menu available to the players involved.

Our design makes it possible to collect very rich data about individual behavior including data about reciprocity. In the next section we explain how our data can be used in relation to various models of social preferences.

[6] This is also related to what Sen (1997) called ‘menu-dependence’. This term refers to the fact that preferences over an outcome may depend on the payoffs at other possible but unreached outcomes. Menu-dependence is not the only aspect of the circumstances around the actual choice set that may have a bearing on how people decide. Sen (1997) refers also to ‘chooser-dependence’: A person’s evaluation of an outcome may depend on the identity or some characteristics of the chooser, i.e. the decision-maker that led to that outcome. The results in Blount (1995) and Charness (2004) are examples of what can be interpreted in terms of chooser-dependence.

[7] In the WPR game, player 1’s payoff loss at the outside option is smaller than in the SPR game, so that *player 2*’s decision to forego the outside option could be considered less kind in WPR than in SPR. However, at the same time, player 2’s payoff is larger at the WPR outside option than at the SPR one, so that player 1 is “more behind” in the latter case and this element may also affect the way in which the foregoing of the outside option is judged. Something similar can be said about the comparison of the SNR and WNR games.

Table 3 Response games

Game SPR	Player 2 chooses (0,1000) or lets player 1 make the choices in Table 1
Game WPR	Player 2 chooses (0,1500) or lets player 1 make the choices in Table 2
Game SNR	Player 2 chooses (2000,1000) or lets player 1 make choices in Table 1
Game WNR	Player 2 chooses (2000,1500) or lets player 1 make choices in Table 2

3 Theoretical background

In this section we briefly discuss the kinds of social preferences that can be potentially relevant in our context. Among the models designed to capture other-regarding preferences two prominent classes can be distinguished: models that only take into account distributional concerns, and models that include other motivational forces. Charness and Rabin (2002) present a simple conceptual model of social preferences in two-person games which embed the different models of preferences in terms of different parameter ranges. Letting x_1 and x_2 be player 1's and player 2's monetary payoffs, the Charness-Rabin utility function of player 1 for choosing action k (A or B) can be written as:

$$\mathrm{U}_{1\mathrm{k}}\,(x_{1k}, x_{2k}) \equiv (1 - \rho\,\mathrm{r}\; - \sigma\,\mathrm{s}\; - \theta\,\mathrm{q})x_{1k} + (\rho\,\mathrm{r}\; + \sigma\,\mathrm{s}\; + \theta\,\mathrm{q})x_{2k} \qquad (1)$$

where $\mathrm{x}_{1\mathrm{k}}$ and $\mathrm{x}_{2\mathrm{k}}$ are the payoffs to player 1 and player 2 associated with the action k.

$$\begin{aligned}\mathrm{r} &= 1 \text{ if } x_1 > x_2, \text{ and r } = 0 \text{ otherwise;}\\ \mathrm{s} &= 1 \text{ if } x_1 < x_2, \text{ and s } = 0 \text{ otherwise;}\end{aligned}$$

$\mathrm{q} = -1$ if player 2 selected an action that made player 1 worse off, and $\mathrm{q} = 0$ otherwise.

In words, player 1's utility is a weighted sum of her own monetary payoff and player 2's monetary payoff, where the weight player 1 places on player 2's payoff may depend on whether player 2 is getting a higher or lower payoff than player 1 and on how player 1 has behaved. The parameters ρ, σ and θ capture various aspects of other-regarding preferences and reciprocal behavior; in all purely distributional models $\theta = 0$. In the Charness-Rabin formulation, the reciprocity element is conceived to only come into play negatively, but one can modify this simply by considering that q = +1 if player 2 has previously behaved in a way that did not hurt player 1. In our case misbehavior or favorable behavior can come into play through menu-dependence, as explained below.

Assuming that each subject maximizes the utility function indicated in Eq. (1), we can obtain information about the parameter values of the utility function through choices between A and B. For instance, if in the choice task 1 (see Table 1), player 1 is indifferent between A and B, we can assume that for this player $\mathrm{U}_{1\mathrm{A}}(x_{1A}, x_{2A}) \equiv (1 - \rho)1000 + \rho\,600 = (1 - \rho)1000 + \rho\,1000 \equiv \mathrm{U}_{1\mathrm{B}}(x_{1B}, x_{2B})$. This equality holds if and only if the parameter $\rho = 0.2$. Consequently, if B is preferred to A then we assume that this player has a parameter $\rho > 0.2$ and *vice versa*. Following with this example, if player 1 prefers B in the first two choice tasks and A in the following ones, then we can assume that this player has a parameter value between $0.25 < \rho < 0.33$. The last column of Table 1 shows the critical value of ρ that would result in player 1 being indifferent between A and B. Any value of ρ below that number would result in a choice of A for tasks 1–5; any value above that number would result in a choice of A for tasks 7–11. The same analysis applies for the parameter σ in Table 2. The critical value of σ is shown in the last column of Table 2.[8]

Charness and Rabin (2002) discuss three types of simple distributional preferences: The first is the *competitive preference* type. A player of this type would be willing to sacrifice some of his own monetary payoff to lower the payoff of the other player. Thus, his utility weight on the monetary payoff of the other player is negative, $\sigma \leq \rho \leq 0$. It is apparent from that last column that for choice tasks 1–5, a competitive player would never choose B. This is because a choice of B would lower his own payoff while at same time increase the other person's payoff relative to choice A. This may be consistent with a positive weight on the other person's monetary payoff, but not with a negative weight as required by competitive preferences. Similarly, a player consistent with competitive preferences would choose B over A for some of decision tasks 7–11. Indeed, if we observe the switching point from A to B in Table 1, we can narrowly identify the range of values for ρ consistent with that switching point. The same analysis applies to the value of σ in Table 2. The switching point from A to B can narrowly define the range of values for σ consistent with that switching point in Table 2. Any observed choices of A in decision tasks 12–16 are inconsistent with a negative value of σ and thus inconsistent with competitive preferences.

A second prominent class of distributional preferences involves *difference aversion*, as modeled by Fehr and Schmidt (1999).[9] Difference aversion implies $\sigma < 0 < \rho < 1$. As we just discussed, the switching points from A to B in Tables 1 and 2 allow us to identify ranges of values for ρ and σ that are consistent with such choices. If these ranges of values are consistent with negative σ and positive ρ, then we can say that the decision maker has a difference aversion preference. There are choices that would be entirely inconsistent with difference aversion. For example, a preference of B over A in decisions 7–11 would be inconsistent with difference aversion because it would imply sacrificing own payoff to hurt a disadvantaged partner. A preference for A over B in decisions 12–16 is also inconsistent with difference aversion (and as discussed above is also inconsistent with competitive preferences) because it implies sacrificing own payoff to help an advantaged partner.

The third additional type that Charness and Rabin (2002) discuss is a *social-welfare preference*, where $0 < \sigma \leq \rho \leq 1$. Such preferences mean that player 1 will never choose to reduce his own monetary payoff (by choosing B) in order to hurt the other person. Thus, choices of B can be ruled out for decision tasks 6–11 and 17–22 for participants who have social welfare preferences. Depending on the magnitude of ρ and σ, some choices of A may be observed and if they are observed, they allow the researcher to define a narrow range of possible values for σ and ρ. Note that the

[8] In choice tasks 6 and 17, player 1 will prefer A over B for any paremeter values of ρ and σ.

[9] Bolton and Ockenfels (2000) present a related model. For some survey evidence on the importance of relative position see Solnick and Hemenway (1998).

perfect substitutes preferences of Andreoni and Miller (2002) are a special case of social-welfare preferences. In the more general version they are sensitive to relative position and to the return on sacrificing.

4 Experimental procedure

The experiment took place in a major European university. A total of 80 participants in player 1 role made 66 decisions each. 79 of these participants provided usable data. The experiment also involved 160 participants in player 2 roles as explained below. The experiment consisted of two parts which we call treatments: the *distributional treatment* (DT, hereafter), involving the choice situations presented in Tables 1 and 2, and the *reciprocity treatment* (RT, hereafter), involving the four response games presented in Table 3. The experiment began with 80 subjects in player 1 role and 80 passive subjects in player 2 role in two separate rooms. At the end of the distributional treatment, the same 80 subjects in player 1 role were asked, following a surprise restart, to participate in the reciprocity treatment (they received no information about the second treatment at the beginning of the first one) with a fresh group of 80 subjects in player 2 role.

Participants kept their roles during the whole treatment and did not have any additional information except the individual payoffs described by the states A and B. Player 1 made 22 choices between two alternative states, A and B, corresponding to the strong and the weak positions.

The type of design we used raises the issue of possible order effects. To take such effects into account we collected data using four different orderings of our six decision environments. These orderings are shown in Table 4, together with the corresponding orderings of active subjects. Our next section shows that there was no significant effect of the order on our results and briefly elaborates on this.

Subjects first made choices in one of the two environments of the distributional treatment. They received a registration sheet for all eleven choices. Then the decisions appeared in a fixed order on an overhead projector. Subjects were prompted to make their decisions in this order, but they could change any of the decisions at any point. At the end of the eleven choices they were asked to go over their decisions and were told that they could make any changes they wished to make. Things then proceeded in the same way for the second environment of the distributional treatment. We believe that, given that subjects had to make so many decisions, the randomization of the order would have introduced additional noise.

The subjects in player 2 role did not make any kind of decision. All participants in this role knew that they would be paid according to the outcome generated by one of the 22 choices of the corresponding player 1 and that they would be anonymously paired with another participant of the other room, both—outcome and partner—to be selected at random.

In the experiment, 1000 units of lab money = 5 euro. The hand-run treatment took less than 30 minutes and average earnings (included a 3 euro show-up fee) were around 10 euro. Experimental instructions are provided in Appendix 1. To make sure subjects understood the instructions we had them answer a questionnaire after the instructions

Table 4 Ordering of the six different environments

Number of subjects (with active role)	Distributional treatment		Reciprocity treatment[a]			
20	STRONG	WEAK	SPR	WPR	SNR	WNR
20	STRONG	WEAK	SNR	WNR	SPR	WPR
21	WEAK	STRONG	WPR	SPR	WNR	SNR
19	WEAK	STRONG	WNR	SNR	WPR	SPR

[a] W stands for weak position, S stands for strong position. PR stands for positive reciprocity, NR stands for negative reciprocity

had been read aloud to the group and just before the experiment began. Nobody made any mistakes.

A total of 160 subjects took part in reciprocity treatment: 80 in the first-mover role of player 2 and 80 in second-mover role of player 1.[10] Again, each group was seated in one of two different rooms. As already mentioned, this treatment was run just after the DT and the subjects in player 1 role were the same subjects in both treatments while the player 2 subjects were different.

In this treatment involving the four response games we applied what is called the strategy elicitation method, which goes back to Selten (1967). The player 1 subjects made their pair-wise choices between two states conditional on the corresponding player 2 letting them choose. In making these decisions, the player 1 subjects knew that their decisions only mattered if the corresponding player 2 subjects had passed up the outside option.[11] In this treatment the player 1 subjects in the second-mover role made a total 44 sequential choices, distributed in 4 blocks of 11 decisions, between two alternative states, A and B. The player 2 subjects (first-mover role) made 4 choices, one for each response game shown in Table 3, between two possibilities: to choose the outside option or to let player 1 choose.

All participants knew that they would be anonymously paired with another participant of the other room, and that they would be paid according to the outcome generated by themselves in one of the 4 blocks, both—outcome and payoffs—to be selected at random. To make sure subjects understood the RT we had them answer a questionnaire after the instructions were read aloud to the group and just before the experiment began. Again, the explanation was repeated until nobody made a mistake (in this case, this was true almost from the beginning).

5 Results

Our statistical analysis consists of three parts. We first present a descriptive overview of our results. Second, we discuss individual level analysis to understand the impact of the different design variables on the decision to sacrifice own payoff. In this analysis

[10] A player 1 left the experimental room once the distributional treatment finished. When we asked him to participate in a new treatment, he refused. Hence, we have complete data from 80 subjects.

[11] On the use of this method see Brandts and Charness (2000, 2011).

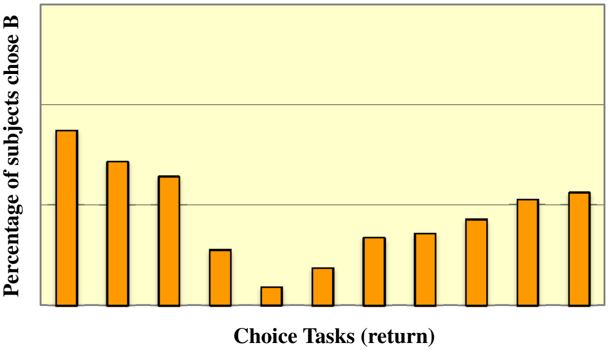

Fig. 1 Return on sacrificing

we include variables to control for a number of design features, e.g., order and time effects. Third, we attempt to uncover utility parameters of the decision parameters as well as characterize individual-level heterogeneity in these parameters. To that end, we present a mixture-model of the distribution of social preferences types in our population, together with an estimation of the preference parameters for each of the types.[12]

5.1 Aggregate level analysis

In this section the effects of different returns on sacrificing on subjects' choices are examined. As mentioned earlier, each binary choice implies a different return on sacrifice; namely, a different impact on the other's payoff for the same amount of sacrifice (one hundred units) of own material payoffs. The set of these returns is the same for the six choice environments that subjects in the player 1role find themselves in. Figure 1 shows aggregate data about the effects of the return on sacrificing, as represented by the percentage of subjects that choose B over A, aggregated over all subjects and all six choice environments.

As Fig. 1 shows, in the aggregate individuals are to a considerable extent willing to sacrifice money to alter the other's payoff, both positively and negatively. In both cases this aggregate willingness depends on what can be considered the natural way on the return on sacrificing money. If one views positive and negative returns with the same absolute value in a symmetric way, then one can state that, in the aggregate, people are more willing to give up money to help than to hurt the other.

We next look at how the effects of the return on sacrificing depend on the influence of reciprocity. Figure 2 shows player 1's aggregate behavior in the reciprocity treatment (PR and NR) together with the behavior, given the same binary choices, in the distributional treatment where player 2 had no option at all. The general features of the effects are the same for all three cases (DT, NR, PR). There are some differences,

[12] We do no study the behavior of player 2 in the response game, since it involves both motivational and strategic aspects and in this paper we are only interested in the former.

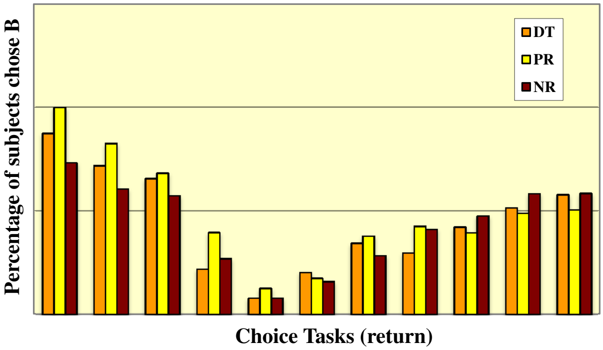

Fig. 2 Return on sacrificing. Reciprocal behavior

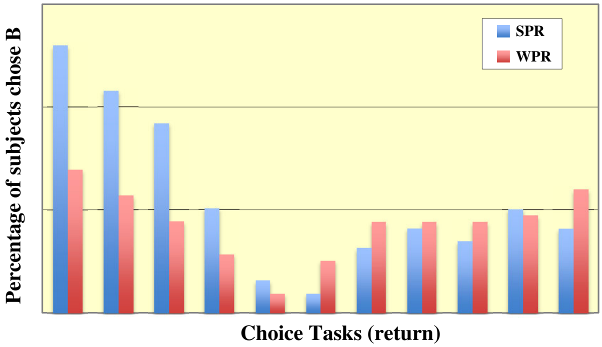

Fig. 3 Return on sacrificing. Strong versus weak. Positive reciprocity treatment

but they appear to be rather secondary ones. For negative returns the differences do not seem to follow any clear pattern. For the four cases of binary choices with positive returns there is a common order of the percentages which are highest for the PR case, intermediate for the NR case and lowest for the DT case. However, note that this is not the pattern that would be consistent with a reciprocity interpretation of our results, which would demand that the percentage of B choices would be lowest for NR, intermediate for DT and highest for PR. At any rate, observe that the differences between the cases are considerably smaller than the respective lowest percentage, a baseline, so that they may not show up as significant in the statistical analysis we present below.

Figures 3 and 4 show the data disaggregated by position. Again, there are some differences but they do not amount to any very clear pattern and it remains to be seen whether they have any statistical validity. This means that the weak evidence for reciprocity that the data presentation of Fig. 3 suggests is not hiding an important interaction effect with the position.

5.2 Individual-level sensitivity to the task

We are interested in the factors that lead player 1 to sacrifice his own monetary payoff—a choice of B. The underlying assumption is that any sacrifice of one’s own monetary

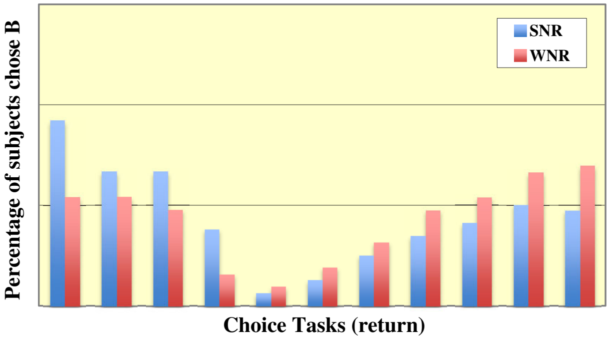

Fig. 4 Return on sacrificing. Strong versus weak. Negative reciprocity treatment

payoff is driven by social preferences, and the ultimate goal of the investigation is to characterize these preferences. The previous section showed the sensitivity of aggregate choices to the design variables. In this section, we report the individual-level regression corresponding to the relationship between the choice of B and the design variables.

The dependent variable is the decision of player 1 in round t (for t = 1,…, 66).[13] The dependent variable takes on the value 1 if player 1 sacrificed his own payoff by choosing B. It is 0 otherwise.

To capture individual differences in responsiveness to design variables, we conduct a panel data random effects probit regression which assumes that the heterogeneity can be captured through the constant parameter in the utility function.[14] The constant parameter is assumed to differ from subject to subject and distributed normally over subjects.

With respect to the independent variables we first describe those that capture the fundamental forces we are interested in analyzing and then move to those that capture potentially important features of our design.

As shown in Tables 1 and 2, our design involves both positive and negative returns. Hence, there are two aspects to player 1 choosing a particular option B: the magnitude of the change in player 2's payoff and the sign of that change. Table 5 shows the results of a random effects Probit estimation. To capture the fact that positive and negative returns correspond to fundamentally different circumstances we introduce the variable *Sacrifice* which captures the sign of the change; it takes the value of 1 when the return on sacrifice is positive and 0 otherwise. *Positive Return* and *Negative Return* are the returns on sacrificing in absolute terms when returns are positive and negative respectively.

Position is a dummy variable that takes the value of 1 for all decisions made from the weak position (and 0 otherwise). *Positive Reciprocity* is a dummy variable set equal to

[13] The 66 rounds correspond to the 11 choices in the six different environments.

[14] In the next section, where we investigate the utility function specification, we are able to divide the players into segments with different sets of preferences. Here, we do not do so and instead opt for random effects because in the absence of a utility specification there is no meaningful economic interpretation to different segments.

Table 5 Random-effects probit estimation

Data	All
Constant	−2.160***
	(0.232)
Round	−0.001
	(0.001)
Order pos	0.321
	(0.249)
Order reciprocity	−0.229
	(0.249)
Sacrifice	0.395***
	(0.123)
Positive return	0.313***
	(0.034)
Negative return	0.198***
	(0.017)
Position	−0.151***
	(0.049)
Positive reciprocity	−0.021
	(0.081)
Negative reciprocity	0.052
	(0.072)
Number of obs	5214 (79 subjects at 66 obs per subject)
Log likelihood	−1858.19

1 for decisions in which the first mover took a positive action and the decision maker is able to positively reciprocate that action (positive return on sacrificing). *Negative Reciprocity* is a dummy variable set equal to 1 for decisions in which the first mover took a negative action and the decision maker is able to negatively reciprocate that action (negative return on sacrificing).

We now describe our procedural variables. Order effects are captured in two different ways. *Order Position* (*Order Reciprocity*) is a dummy variable which controls for one of the order effects and takes the value of 1 for the subjects that first play blocks from the *Strong Position* (blocks with *Positive Reciprocity*) and zero otherwise. Going back to the four sequences shown in Table 4, for the first row both *Order Position* and *Order Reciprocity* have a value of 1, while for the second to fourth rows the values of *Order Position* and *Order Reciprocity* are 1 and 0, 0 and 1 and 0 and 0, respectively. *Round* refers to the order in the sequence, from 1 to 11, in which the binary decisions were made. Table 5 shows the results of our estimation using the entire data set.

The results shown in Table 5 strongly suggest that observed behavior does not depend on the variables we included to control for some of the procedural details of our design. *Round, Order Position* and *Order Reciprocity* are insignificant. The order in which subjects make decisions within each block and the order of the treatments has no apparent effect on behavior.

The results of Table 5 show that subjects strongly react to the nature (helping or hurting) and the returns on their decisions. Their willingness to sacrifice significantly increases whenever they can help others with their sacrifice as *Sacrifice* is positive and strongly significant.

Subjects are shown to be responsive to the returns on their decisions. When the positive impact on others from choosing B (positive return) rises, subjects are more likely to help others. When their negative impact on others from choosing B rises (negative return) they are more likely to hurt others. The slope on the negative return is flatter, suggesting that responsiveness to negative impact is lower. Note that the propensity to hurt others when the negative impact rises should be considered together with the positive coefficient on sacrifice. That is, overall, subjects are far less likely to sacrifice their own payoff for the purpose of hurting others than they are to sacrifice their own payoff for the purpose of helping others. However, conditional on choosing to hurt others, they are more likely to do so when the impact of their action relative to their sacrifice is larger.

Note also that the probability of choosing B is significantly altered when comparing decisions made from the weak and the strong position in all cases. *Position* is negative and significant. This means that subjects are willing to sacrifice more when they make decisions from a position of strength (their earnings are above the payoff of the other player) than from a weak position.

In contrast, reciprocity does not affect the probability of sacrificing. *Positive/Negative Reciprocity*, have no significant effect.

The estimates clearly confirm the informal impression given by the aggregate data shown above. Return, sacrifice and relative position treatment effects are highly significant and have the expected sign, while the reciprocity variables are not significant.

We now know that overall reciprocity (related to menu-dependence) does not matter and that the relative price and the position do matter. However, we have not yet identified more precisely what type of preference models the different individuals' behavior are consistent with. The only way to find out about this is to look more closely at the individual level data. In the next section we present the results of mixturemodel estimation.

5.3 Estimating utility functions

In the previous section, we reported a random effects regression whose purpose was to characterize the relationship between design variables and choice. In that analysis, we handled heterogeneity through a continuous random effects specification because discretizing the parameter space would have had no natural interpretation in the absence of utility specification.

Once a utility function has been specified, one can postulate different types of behavior corresponding to different parts of the utility parameter space. The mixture model is an ideal approach to identify subpopulations of types with different social preferences (e.g., Iriberri and Rey-Biel 2013; Cappelen et al. 2007; Conte and Moffatt 2010). Table 6 shows the results of our mixture-model estimation and Appendix 2 contains details of the estimation.

Table 6 Mixture model estimation, n = 5214 (79 subjects; 66 obs per subject)

Term	2-segments	3 segments	4 segments	5 segments	
% Selfish		0.446*	0.432*	0.405*	
% Zero intelligence	0.304*	0.278*	0.085*	0.113*	
%Social-welfare maximizers	0.696*	0.275*	0.277*	0.277*	
% Inequality-averse				0.116*	
% Competitive			0.206*	0.089*	
ρ 1	0.191*	0.347*	0.345*	0.342*	Social welfare maximizers
σ 1	0.142*	0.267*	0.267*	0.268*	
θ 1	−0.0005	−0.041	−0.04	−0.039	
Constant1	100	100	100	100	
ρ 2				0.243*	Inequity aversion
σ 2				−0.351*	
θ 2				0.032	
Constant2				100	
ρ 3			−0.166*	−0.536*	Competitiveness
σ 3			−0.617*	−0.466*	
θ 3			−0.007	−0.025	
Constant3			100	100	
Epsilon	0.371*	0.030*	0.027*	0.020*	Selfish
Log likelihood	−2007	−1819	−1649	−1574	
# of parameters	6	7	12	17	
AIC	4026	3652	3322	3182	
BIC	4065	3698	3401	3294	

* Significant at 5 % level

The upper part of the table shows the distribution of types and the lower part the parameter estimates for the different types. In our estimation we impose the existence of a selfish type as well as of a zero-intelligence who chooses all actions with equal probabilities, but we do not impose any restrictions on the parameters of the types with social preferences. All types with the corresponding specific values of the parameters emerge endogenously.

We assess the models with 2–5 segments on log likelihood, AIC, and BIC.[15] We find that the five-segment model performs best on all three measures.

Focusing on the results for the model with five segments, we start with the fractions for the two restricted types. The zero-intelligence or noisy traders amount to only 11.3 %. This means that we are able to classify[16] almost 90 % of our subjects into

[15] AIC stands for Akaike Information Criterion and is defined as $AIC = -2Loglikelihood + 2k$. BIC stands for Bayesian Information Criterion and is defined as $BIC = -2LogLikelihood + kln(n)$. For both criteria, lower values are preferred to higher values.

[16] The proportions of subjects from each type are generated as estimates of the mixture model, without the need to classify any individual subjects. Nevertheless, individual subjects can indeed be classified and this

preference-based types. We can see that the *selfish* type, $\sigma = 0$ and $\rho = 0$, is the most frequent one with 40.5 % of the subjects.

The next more frequent type—the first endogenously emerged one— is the *social-welfare maximizer*, since the estimated parameter values satisfy $0 < \sigma \leq \rho \leq 1$. Note that both σ and ρ are significantly different form zero, with ρ being significantly larger than σ (restricting the two to be equal yields LL = 1578.7 versus 1573.9, Likelihood Ratio test statistic$= 4.8 \times 2 = 9.6$, distributed Chi-square with 1 d.f, yielding p-value< 0.01). θ is insignificant. The *social-welfare maximizer* type covers 27.7 % of subjects and is the most frequent social preferences type.

There are two other social preferences types emerging from the data. *Inequity-averse* subjects, characterized by $\sigma < 0 < \rho < 1$, account for 11.6 % of participants and *competitive* subjects cover 8.9 %. In both cases the estimates satisfy the theoretical restriction. For the *inequity-averse* type ρ is positive and significantly larger than the negative coefficient for σ. For the competitive type, ρ and σ are negative and not significantly different from each other (LR test statistic = $(1574.4 - 1573.9) * 2 = 1$, chi-square d.f.=1, *p* value = 0.32).

We now look more carefully at the estimated parameter values. Note first that, that all estimated parameter configurations involve an insignificant estimate of θ. This is consistent with the results from the random effects estimation discussed earlier. Menu-dependent reciprocity is absent from our data.

All the other parameter value estimates of the different types are consistent with the theoretical discussion in Sect. 3. *Social-welfare maximizers* do trade-off their own payoff with that of the other, but assign higher weight to selfish-payoff than to the other's payoff– in contrast to what pure substitution would imply (a regression with rho1 fixed at 0.5 yields a LL = 1596.6 which by the chi squared test gives a *p* value< 0.01) and more so if behind in payoff.

The parameter for the *inequality-averse* type exhibits a larger absolute aversion to being behind than to being ahead. The parameter values for what we call the *competitive* type equally competitive being ahead or behind.

6 Discussion and conclusions

We present a new experimental design that uses dictator games with a manipulation of the relative position of the dictator and the prior action by the recipient to generates information that directly pertains to *individuals'* social preferences. Dictator games remain the workhorse in experimental economics for mapping social preferences (though challenges have been raised[17]). In that respect our design provides both

Footnote 16 continued
classification roughly matches the proportions estimated. Individual classification is done on the posteriors. A subject is classified as a particular type if the posterior probability of him belonging to that type is higher than the corresponding probability of him belonging to any other type of the types specified.

[17] List (2007) shows that changing the framing of the dictator game results in different behavior. Based on these results, he challenges the notion that dictator games reveal social preferences on the grounds that social norms present a major confound. Cooper and Kagel (2011) and Fehr and Schmidt (2006) provide similar questions of robustness in dictator games and review challenges.

the ability to derive more precise identification of parameter values and the ability to separate out competing explanations for dictator game behavior,

Our data present the following patterns: First, a large portion of observed behavior cannot be understood in purely individualistic terms since subjects' behavior is sensitive to the cost of helping and hurting others. Second, we find strong evidence that the weight on others' payoffs depends on whether the decision maker is in an advantageous or disadvantageous position.

Third, it is noteworthy that we find no evidence of reciprocity in the data. Reciprocity remains one of the most elusive preferences to demonstrate conclusively in the laboratory and existing evidence for reciprocity is quite mixed. Brandts and Solà (2001), Falk et al. (2003) and Cox (2004; see Cox et al. 2008 for axiomatic characterization of the data) find favorable evidence for reciprocity, while Charness and Rabin (2002) Bolton et al. (1998), Bolton et al. (2000) and Cox and Deck (2005) do not find it. In games such as the ultimatum and trust games which have been thought to conclusively demonstrate negative and positive reciprocity, respectively, the reciprocity explanation and its robustness are increasingly challenged.[18]

Given the many explanations proposed in the literature, we can only conjecture on why it reciprocity is not showing in our data. One possible conjecture is called the "complicity effect" (Charness and Rabin 2002), which loosely means that the mere action by player 2 alleviates the responsibility player 1 takes for the outcome. In our setting, Player 2 subjects who opted to give player 1 the choice gave up in potential payoff far greater than any payoff they could possibly earn by giving player 1 the choice. Thus, player 1 could surmise that player 2 values the monetary payoff to player 1 a great deal. This in turn may justify a more selfish action on the part of player 1.

Another possible explanation may pertain to the response-elicitation method. It is possible that in certain environments direct elicitation invokes more reciprocal response. That said, there is no consensus in the literature that there is a "right" approach. Indeed, there are relevant economic situations in which decision makers must form contingent strategies in advance. In addition, the survey of direct comparisons of the two elicitation methods contained in Brandts and Charness (2011) finds that there are more studies that do not find a difference than studies that find one. In our particular case, collecting the kind of individual data required with the direct response method is not feasible. Specifically, using the direct elicitation method would not allow us to collect a complete decision set for every participant. Missing these important parts of the reaction space would not permit the type of analysis we report here.

Our design makes it possible to go one step beyond in characterizing of individual preferences by estimating the parameter values of their utility functions. The results of our mixture-model estimation show that observed non-individualistic behavior is very heterogeneous across individuals. As also stressed by Andreoni and Miller (2002) people conform to more than one model of social preferences. In relation to the debate

[18] The best known evidence on negative reciprocity comes from the ultimatum game. However, there is a great deal of literature challenging the robustness of these results (Falk et al. 2003; Brandts and Solà 2001; Stahl and Haruvy 2008). The interpretation of trust game results as evidence for positive reciprocity has been challenged as well (e.g., Cox and Deck 2005).

about what kind of other-regarding preferences are more effective in explaining behavior, our results are to some extent compatible with those in Charness and Rabin (2002). They also concord partially with the findings of Engelmann and Strobel (2004), who study social preferences in the context of three-person games and do not investigate any issues related to reciprocity. Like them, we find that the influence of social welfare preferences is stronger than that of inequality aversion. In our data social welfare maximizers are not of the perfect substitutes type; subjects do give more weight to own payoff than to that of others.

Acknowledgments Financial support by the Spanish Ministry of Economics and Competitiveness (Grant: ECO2011-29847-C02-01), the *Generalitat de Catalunya* (Grant: 2009 SGR 820), the Barcelona GSE Research Network, and CONSOLIDER-INGENIO 2010 (CSD2006-00016) is gratefully acknowledged. We thank Juan Mañez for helpful comments on an earlier version. Remaining errors are ours.

Appendix 1: Instructions

Instructions distributional treatment

This is an experiment about decision making. You will be paid for participating, and the amount of money you will earn depends on the decisions that you and the other participants make. At the end of the experiment you will be paid privately and in cash for your decisions. You will never be asked to reveal your identity to anyone during the course of the experiment.

In this experiment there are two types of subjects, x and y. As subject x will make 22 sequential choices in 2 blocks of 11 decisions between two alternative states (A and B). Each decision is independent from each your other decisions. Your payoffs in the experiment depend on your decisions. You will be anonymously paired with a subject y.

As subject y will not make any kind of decision. You will be anonymously paired with a subject x and your payoffs in the experiment depend on subject x's choices.

To make decisions you only have to circle in the control sheet one of the two options A and B in each round.

At the end of the experiment you will thus have 22 outcomes from the rounds played, only one of these outcomes will be selected for payoffs.

Instructions reciprocity treatment

This is an experiment about decision making. You will be paid for participating, and the amount of money you will earn depends on the decisions that you and the other participants make. At the end of the experiment you will be paid privately and in cash for your decisions. You will never be asked to reveal your identity to anyone during the course of the experiment.

In this experiment there are two types of subjects, x and z. As subject x will make 44 sequential choices in 4 blocks of 11 decisions between two alternative states (A and B). Each decision is independent from each your other decisions. You will be anonymously paired with a subject z. As player x your decisions will only affect the payoffs if player z opts to give you the choice.

As subject z will be anonymously paired with a subject x and will make 4 decisions, one per block, between two possibilities: to choose or to let player x choose. The player x knows that her decisions will only affect the payoffs if subject z opts to give her the choice.

To make decisions you only have to circle in the control sheet one of the two options in each block.

At the end of the experiment you will be paid according to the outcomes generated by yourselves in the 4 blocks, only one of these outcomes will be selected for payoffs.

Appendix 2: Mixture model estimation

A utility-based behavioral type is defined by two equations. The first equation is a utility function, generally of the form of Eq. (1), repeated below for convenience. The second equation identifying a behavioral type maps the utilities of the different actions in the game to a probability distribution over these actions. Of the five behavioral types we discussed, the first three types correspond to the utility specification of equation (1).

$$\mathrm{U_{ik}} \equiv (1 - \rho\,\mathrm{r} \;- \sigma\,\mathrm{s} \;- \theta\,\mathrm{q})\;\; \mathrm{x_{ik}} + (\rho\,\mathrm{r} \;+ \sigma\,\mathrm{s} \;+ \theta\,\mathrm{q})\;\; \mathrm{x_{jk}} \quad (1)$$

where $\mathrm{x_{ik}}$ and $\mathrm{x_{jk}}$ are the payoffs to player 1 and player 2 associated with the action k.

Equation (1) can be alternatively expressed as

$$\mathrm{U_{ik}} \equiv \mathrm{x_{1k}} - (\rho\,\mathrm{r} \;+ \sigma\,\mathrm{s} \;+ \theta\,\mathrm{q})\,(\mathrm{x_{1k}} - \;\mathrm{x_{2k}}) \quad (1')$$

The relevant construct for the probability mapping is the difference in utilities between the two actions. This is computed as:

$$\Delta\mathrm{U_1} \equiv \Delta\mathrm{x_1} - (\rho\,\mathrm{r} \;+ \sigma\,\mathrm{s} \;+ \theta\,\mathrm{q})\,(\Delta\mathrm{x_1} - \Delta\mathrm{x_2}) \quad (1'')$$

where Δxi is the difference between player 1's payoffs to choosing A and B, and similarly for player 2. Note, however, that by design $\Delta\mathrm{x_1} = 100$. Hence

$$\Delta\mathrm{U_1} \equiv 100 + (\rho\,\mathrm{r} \;+ \sigma\,\mathrm{s} \;+ \theta\,\mathrm{q})\,(\Delta\mathrm{x_2} - 100) \quad (1''')$$

The probability mapping from utility to choice is the logit:

$$\Pr(ChoiceA) = \frac{\exp\left(\frac{1}{\lambda}U_A\right)}{\exp\left(\frac{1}{\lambda}U_A\right) + \exp(\frac{1}{\lambda}U_B)} \qquad (2)$$

This equation adds a scaling parameter λ to the estimation.

Before we proceed to discuss the mixture model, a few more notes on the non-utility based types. There are two types that are not utility-based. The first is the level-0 which is a type that chooses both actions with equal probability of 0.5. Without this type, the scaling parameter on the utility-based types tends to pick up noisy behavior, possibly leading to erroneous inferences about utility weights. The second is the selfish type. The selfish type is critical because it is the default type in economic theory. Errors by the selfish type occur with probability epsilon, which is being estimated.

Mixture model estimation is based on the notion that there are discrete latent subpopulations of players behaving in distinct ways. The method first requires estimating a likelihood function for each player's joint choices conditional on being of particular type m. For each of the 79 players in the data, indexed i = 1,…, 79, there are 66 choices, indexed by d = 1…66.

The likelihood for subject i's combined 66 choices condition on being type m is expressed as:

$$L_{i,m}^{cond} = \text{L(subject i's joint choices | type m)} = \prod_{d=1}^{66} P(choice_{id}|type\, m)$$

The unconditional likelihood for subject i's combined 66 choices condition on being type m can be computed as:

$$L_{i}^{uncond} = \text{L(subject i′s joint choices)} = \sum_{m=1}^{M} \alpha_m L_{i,m}^{cond}$$

The α_m parameter is the unconditional probability of any player being of type m. It can also be interpreted as the proportion of m type players in the population.

It is important to stress that the three utility-based types are ex-ante identical in specification. The separation into type occurs in an unrestricted manner. That is, we did not restrict the parameter space to the positive quadrant for the first type, etc. The types simply emerged unrestricted from the data.

References

Almas I, Cappelen A, Sørensen E, Tungodden B (2010) Fairness and the development of inequality acceptance. Science 328:1176–1178

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

Andreoni J, Harbaugh W, Vesterlund L (2008) Altruism in experiments. The new palgrave dictionary in economics, 2nd edn. Macmillan, New York

Becker A (2013) Accountability and the fairness bias: the effects of effort vs. luck. Social Choice and Welfare 41:685–699

Blanco M, Engelmann D, Normann H (2011) A within-subject analysis of other-regarding preferences. Games Econ Behav 72:321–338

Blount S (1995) When social outcomes aren't fair: the effect of casual attributions on preferences. Organ Behav Hum Decision Process 63:131–144

Bolton G, Ockenfels A (2000) ERC: a theory of equity, reciprocity, and competition. Am Econ Rev 90:166–193

Bolton G, Brandts J, Ockenfels A (1998) Measuring motivations for the reciprocal responses observed in a simple dilemma game. Exp Econ 1:207–219

Bolton G, Brandts J, Katok E (2000) How strategy-sensitive are contributions? A test of six hypotheses in a two-person dilemma game. Econ Theory 15:367–387

Brandts J, Charness G (2000) Hot vs. cold: sequential responses in simple experimental games. Exp Econ 2:227–238

Brandts J, Charness G (2011) The strategy versus the direct-response method: a first survey of experimental comparisons. Exp Econ 14:375–398

Brandts J, Schram A (2001) Cooperation and noise in public goods experiments: applying the contribution function approach. J Public Econ 79:399–427

Brandts J, Solà C (2001) Reference points and negative reciprocity in simple sequential games. Games Econ Behav 36:138–157

Cappelen A, Hole A, Sorensen E, Tungodden B (2007) The pluralism of fairness ideals: an experimental approach. Am Econ Rev 97:818–827

Cappelen A, Hole A, Sorensen E, Tungodden B (2011) The importance of moral reflection and self-reported data in a dictator game with production. Social Choice Welf 36:105–120

Charness G (2004) Attribution and reciprocity in an experimental labor market. J Labor Econ 22:665–688

Charness G, Rabin M (2002) Understanding social preferences with simple tests. Q J Econ 117:817–869

Charness G, Haruvy E, Sonsino D (2007) Social distance and reciprocity: an internet experiment. J Econ Behav Organ 63:88–103

Conte A, Moffatt P (2010) The econometric modeling of social preferences. Working paper. Jena Research Papers 2010–042

Cooper D, Kagel J (2011) Other regarding preferences: a selective survey of experimental eesults. In: Kagel J, Roth A (eds) The handbook of experimental economics. Princton University Press, Amsterdam

Cox J (2004) How to identify trust and reciprocity. Games Econ Behav 46:260–281

Cox J, Deck C (2005) On the nature of reciprocal motives. Econ Inq 43:623–635

Cox J, Friedman D, Sadiraj V (2008) Revealed altruism. Econometrica 76:31–69

Dufwenberg M, Kirchsteiger G (2004) A theory of sequential reciprocity. Games Econ Behav 47:268–298

Engelmann D, Strobel M (2004) Inequality aversion, efficiency and maximin preferences in simple distribution experiments. Am Econ Rev 94:857–869

Falk A, Fehr E, Fischbacher U (2003) On the nature of fair behavior. Econ Inq 41:20–26

Falk A, Fischbacher U (2006) A theory of reciprocity. Games Econ Behav 54:293–315

Fehr E, Schmidt K (1999) A theory of fairness, competition, and cooperation. Q J Econ 114:817–868

Fehr S, Schmidt K (2006) The economics of fairness, reciprocity and altruism—experimental evidence and new theories. In: Kolm S, Ythier J (eds) Handbook on the economics of giving, reciprocity and altruisn. North-Holland, Amsterdam

Fisman R, Kariv S, Markovits D (2007) Individual preferences for giving. Am Econ Rev 97:1858–1876

Iriberri N, Rey-Biel P (2013) Elicited beliefs and social information in modified dictator games: what do dictators believe other dictators do? Quant Econ 4:515–547

Johansson-Stenman O, Carlsson F, Daruvala D (2002) Measuring future grandparents preferences for equality and relative standing. Econ J 112:362–383

Konow J (2009) Is fairness in the eye of the beholder? An impartial spectator analysis of justice. Social Choice Welf 33:101–127

List J (2007) On the interpretation of giving in dictator games. J Polit Econ 115:482–493

Offerman T, Sonnemans J, Schram A (1996) Value orientations, expectations, and voluntary contributions in public goods. Econ J 106:817–845

Rabin M (1993) Incorporating fairness into economics and game theory. Am Econ Rev 83:1281–1302

Selten R (1967) Die Strategymethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In: Beiträge zur Experimentellen Wirtschaftsforschung. Sauermann H (ed). Mohr, Tubingen p 136–168

Sen A (1997) Maximization and the act of choice. Econometrica 65:745–779

Solnick S, Hemenway D (1998) Is more always better?: a survey of positional concerns. J Econ Behav Organ 37:373–383

Stahl D, Haruvy E (2008) Subgame perfection in ultimatum bargaining trees. Games Econ Behav 63:292–307

Zizzo D, Oswald A (2001) Are people willing to pay to reduce others' incomes? Annalesd'Economie et de Statistique 63–64:39–62