DEPARTMENT OF ECONOMICS

ISSN 1441-5429

DISCUSSION PAPER 26/11

Testing the Analytical Framework of Other-Regarding Preferences

M. Vittoria Levati,[1 2] Aaron Nicholas,[3] and Birendra Rai[4]*

Abstract

This paper aims to assess the empirical validity of the overall theoretical framework of other-regarding preferences. We focus on those preference axioms that are common to all the prominent theories of outcome-based other-regarding preferences. This common set of preference axioms leads to a testable implication: the strict preference ranking of self over a _nite number of alternatives lying on any straight line in the space of material payoffs to self and other will be single-peaked. We elicit the strict preference rankings of experimental subjects in variants of dictator and trust games using a mechanism that induces truthful revelation under quite weak assumptions. The data allow us to document the extent of single-peakedness and identify who violates single-peakedness. Potential reasons for violations of single-peakedness are delineated and the implications of our findings for theoretical modeling of other-regarding preferences are discussed.

Keywords: Other-regarding preferences, expected utility theory, single-peaked preferences, Experiments

Acknowledgement We thank Klaus Abbink, Simon Angus, Tim Cason, Nick Feltovich, Lata Gangadharan, Ben Greiner, Werner Guth, Glenn Harrison, Vai-Lam Mui, Nikos Nikiforakis,Robert Slonim, Tom Wilkening, Roi Zultan, seminar participants at Monash University and University of Melbourne, and participants at the Sixth Australia New Zealand Workshop on Experimental Economics for helpful suggestions. Christoph Goring, Andreas Matzke, and Claudia Zellmann provided valuable assistance in conducting the experiments.Financial support from Monash University and the Max Planck Institute of Economics is gratefully acknowledged. The usual disclaimer applies.

[1] Max Planck Institute of Economics, Jena 07745, Germany
[2] Department of Economics, University of Verona, Verona 37129, Italy
[3] Graduate School of Business, Deakin University, Burwood 3125, Australia
[4] Department of Economics, Monash University, Clayton 3800, Australia
*To whom correspondence should be addressed. E-mail: birendra.rai@monash.edu

1. Introduction

Several models have been proposed to account for the empirical finding that individuals exhibit other-regarding behavior. Most of the prominent models of outcome-based other-regarding preferences (including, Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002) can be subsumed under a common theoretical framework. Cox, Friedman, and Sadiraj (2008) highlight that outcome-based models of other-regarding preferences share a set of *core* preference axioms.

Preferences are modeled with the help of *a* binary relation defined over a set whose elements specify the material payoffs for *self* and *other*. They are assumed to be transitive, complete, continuous, (strictly or weakly) convex, and strictly monotonic with respect to payoff to *self*. Preferences are allowed to be non-monotonic with respect to the payoff of *other* by some models but not by others. Non-monotonicity helps capture the intuitive idea that an individual might be willing to sacrifice his own payoff in order to decrease the payoff of another individual. The indifference curves of an individual can thus be positively sloped at certain points in the space of payoffs to *self* and *other*.

A testable implication follows from the core axioms regarding the structure of preference rankings. Intuitively, the strict ordinal preference ranking of an individual over a finite number of alternatives lying on any straight line in the space of payoffs to *self* and *other* will be *single-peaked*, where the term 'single-peaked' is being used precisely in the way it is used in social choice theory (Black, 1958).

This paper aims to assess the empirical validity of the assumptions underlying the whole class of outcome-based models of other-regarding preferences by eliciting preference rankings and examining the extent of single-peakedness. Self-regarding preferences are treated as a special case of other-regarding preferences in this framework. Consequently, preferences of self-regarding subjects will also be single-peaked if different alternatives specify different payoffs for *self*.

Figure 1 illustrates the idea of single-peakedness by considering five alternatives lying on a downward sloping straight line in the two-dimensional space of material payoffs to *self* and *other*. Suppose the preferences of *self* can be modeled with the help of *a* binary relation which is complete, transitive, continuity, strictly convex, monotonic with respect to own payoff but not necessarily monotonic with respect to the payoff of *other*. Figure 1(A) illustrates a family of indifference curves for *self* that are consistent with these preference axioms. Given the indifference curves, we can identify the ordinal preference ranking of the five alternatives lying on the downward sloping line.[1]

[1] Any indifference curve will contain none, one, or at most two out of the five alternatives on the downward sloping line if the preference axioms are satisfied. For ease of exposition we are illustrating the case where all the five alternatives lie of distinct indifference curves. Section 2 provides details of the case where two alternatives may lie on the same indifference curve.

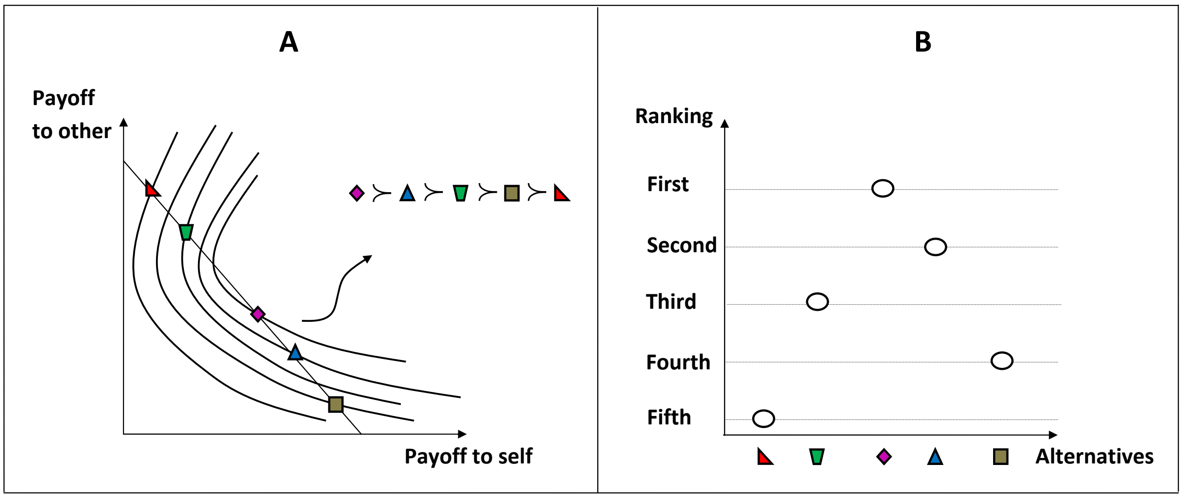


Figure 1: Core axioms imply single-peakedness

We can order the five alternatives on a one-dimensional axis with respect to the payoff to *self* (or, *other*) even though each alternative lies in a two-dimensional space. The five alternatives are ordered on the horizontal-axis in terms of increasing payoff to *self* as shown in Figure 1(B). The ordinal preference ranking over the five alternatives is denoted on the vertical-axis. Intuitively, preference rankings over a set of alternatives that can be ordered on a one-dimensional axis are single-peaked if, as we move from left to right along the axis on which the alternatives are ordered, then (i) preferences monotonically increase, or (ii) monotonically decrease, or (iii) first they monotonically increase and then monotonically decrease. It can be easily seen that the ordinal preference ranking of *self* over these alternatives is single-peaked. as case (iii) holds in Figure 1(B).

All the questions addressed in this paper revolve around the idea of single-peakedness. Section 2 provides a simple example to clarify the nature and usefulness of the questions we shall investigate. We also state a formal proposition linking the core preference axioms in models of other-regarding preferences with single-peakedness of preference rankings.[2]. Section 3 describes the experimental design and procedures including a detailed description of the mechanism used to elicit rankings. Section 4 lays out the main hypotheses and Section 5 presents the results. Section 6 concludes with a detailed discussion of the results with an emphasis on delineating the potential reasons for non-single-peakedness, their relation to the existing literature, and avenues for future work.

[2]To the best of our knowledge, Andreoni, Castillo, and Petrie (2003) is the only prior study that mentions the notion of single-peakedness in the context of other-regarding preferences. However, our study differs from theirs. We exploit the idea that preference rankings *rankings* of dictators over a fixed set of alternatives will be single-peaked. In contrast, Andreoni et al. (2003) exploit the idea that the locus of *choices* of responders in a series of suitably constructed ultimatum games will be single-peaked.

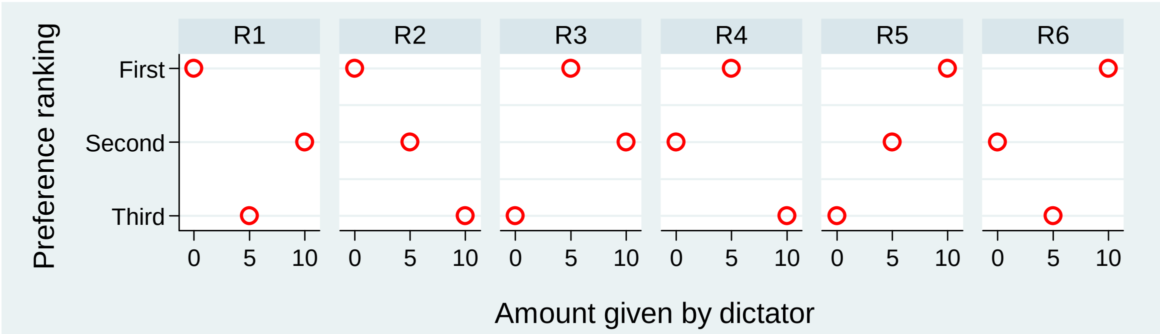


Figure 2: The six possible preference rankings in the mini-dictator game D

2. Single-peakedness

Consider a mini-dictator game D involving two agents – the inactive recipient and the active dictator who can give 0, or 5, or 10 out of 20 to the recipient. It is standard practice to label any subject who would give 10 as an 'other-regarding' agent and claim that his behavior can be accommodated by at least one of the models within the framework of Cox et al. (2008).

This claim amounts to making the implicit assumption that the subject's (strict) ranking over the three alternatives is giving $10 \succ 5 \succ 0$ rather than $10 \succ 0 \succ 5$. However, we can not rule out the latter ranking solely on the basis of the subject's observed *choice* of giving 10.[3] To the best of our knowledge, the latter ranking can not be accommodated in any existing outcome-based or intention-based theory of other-regarding preferences.

Figure 2 illustrates the six possible strict rankings of the three alternatives in game D. Rankings R2 to R5 are single-peaked. Rankings $R2$ and $R5$ reflect monotonically decreasing and monotonically increasing preferences, respectively. $R3$ and $R4$ reflect preferences that first increase and then decrease as we move along the horizontal axis.

Rankings $R1 \equiv 0 \succ 10 \succ 5$ and $R6 \equiv 10 \succ 0 \succ 5$ are incompatible with the existing framework of other-regarding preferences. These two rankings are *non-single-peaked* since preferences first decrease and then increase as we move along the horizontal axis. A subject whose ranking is $R1 \equiv 0 \succ 10 \succ 5$ starts with the most selfish choice of giving 0 as his first-ranked alternative but then his *selfishness runs out* and he ranks giving 10 above giving 5. A subject whose ranking is $R6 \equiv 10 \succ 0 \succ 5$ starts with the most generous choice of giving 10 but then his *generosity runs out* and he ranks giving 0 above giving 5. Rankings $R1$ and $R6$ thus violate single-peakedness in different ways and lead to different behavioral interpretations.

[3]We ignore the possibility of indifference in the present discussion in order to clearly convey our basic ideas and discuss the reasons for doing so at the end of Section 2.1.

It is possible that the underlying preferences are single-peaked for most of the subjects who choose to give 0 but non-single-peaked for most of the subjects who give 10. In other words, generosity may run out significantly more often than selfishness. If this is indeed the case, then the framework of other-regarding preferences would fail to accommodate those very subjects whom it essentially aims to accommodate (i.e., the 'non-neocalssical' subjects who give something other than 0). In contrast, our confidence in the empirical validity of the framework would be greatly enhanced if there is no systematic difference in single-peakedness of the underlying preference rankings across subjects who differ in their choices and the overall extent of non-single-peakedness is quite low. Our experiments are therefore designed to measure the *extent* of single-peakedness and identify *who* violates single-peakedness by directly eliciting the preference rankings of subjects.

2.1. Core axioms imply single-peakedness

Let $\succsim$ be the binary preference relation of an agent defined over the set $\mathbb{X} \subset \mathbb{R}^2_+$. A representative element of $\mathbb{X}$ will be denoted as (x^s, x^o), where $x^s \in \mathbb{R}_+$ denotes the material payoff to *self* and $x^o \in \mathbb{R}_+$ denotes the material payoff to *other*. Let $\succ$ and $\sim$ be the strict preference relation and the indifference relation derived from $\succsim$.

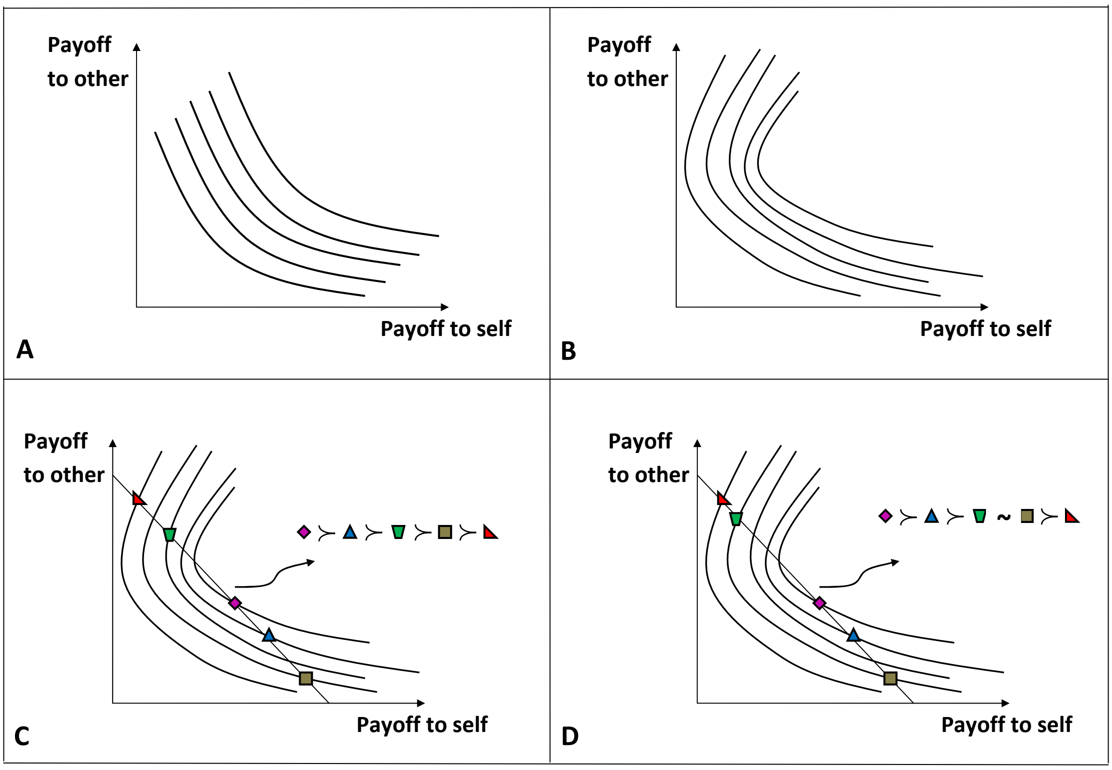
Let $\succsim$ be complete, transitive, continuous, strictly convex, and strictly monotonic with respect to x^s. Completeness, transitivity, and continuity of $\succsim$ imply there exists a real-valued utility function $u : \mathbb{X} \to \mathbb{R}$ which represents $\succsim$ and strict convexity implies that indifference curves in the payoff space can not have linear segments. If $\succsim$ is strictly monotonic in both x^s and x^o, then indifference curves in the payoff space must be negatively sloped everywhere as shown in Fig. 3(A). However, if $\succsim$ is strictly monotonic in x^s but not always strictly monotonic in x^o, then an indifference curve can be positively sloped at certain points in the payoff space as shown in Fig. 3(B).

Let $\mathbb{A} = \{a_1, a_2, \ldots, a_n\}$ be a finite set of alternatives lying on any strictly downward sloping or strictly upward sloping straight line in the space of payoffs to *self* and *other*.

Proposition 1: *Any strict ordinal ranking consistent with a weak (or strict) ordinal ranking over elements of $\mathbb{A}$ will be single-peaked if preferences are complete, transitive, continuous, strictly convex, and strictly monotonic in x^s.*

The proof of this proposition is quite straightforward and we provide a brief sketch. Consider any downward sloping straight line in the space of payoffs. Given the assumptions on $\succsim$, any alternative $a_i \in \mathbb{A}$ must lie on one, and only one, indifference curve. However, an indifference curve may pass through none, one, or at most two alternatives belonging to $\mathbb{A}$ as shown in Figures 3(C) and 3(D).

Figure 3: (A) Indifference curves corresponding to a preference relation that satisfies the core axioms and is also monotonic with respect to payoff of *other*. (B) Indifference curves corresponding to a preference relation that satisfies the core axioms but is not always monotonic with respect to payoff of *other*. (C) An example where no indifference curve contains more than one of the alternatives in a finite set. (D) An example where an indifference curve contains two of the alternatives in a finite set.



We can order the alternatives in the set $\mathbb{A}$ on a horizontal one-dimensional axis in terms of the payoff to *self* (or *other*) and plot their ordinal ranks on the vertical axis. If all alternatives lie on different indifference curves, then they can easily be assigned different ordinal ranks and the obtained ranking over the alternatives will automatically be a strict ranking. Single-peakedness requires that, among any two alternatives that are both on the left or the right of the first-ranked alternative, the alternative that is closer to the first-ranked alternative is ranked above the alternative that is farther from the first-ranked alternative. It is easily verified that the strict ordinal ranking will be single-peaked.

We will get a weak ordinal ranking over the alternatives when two alternatives in $\mathbb{A}$ lie on the same indifference curve. Suppose the obtained weak ordering is $a_3 \succ a_2 \succ a_1 \sim a_4 \succ a_5$. Two strict rankings are consistent with this weak ranking: $a_3 \succ a_2 \succ a_1 \succ a_4 \succ a_5$ and $a_3 \succ a_2 \succ a_4 \succ a_1 \succ a_5$. Both of these strict rankings will be single-peaked.[4]

[4]We omit the explanation for the case where the finite number of alternatives lie on a strictly upward sloping straight line since it is similar to the case of a strictly downward sloping line.

Single-peakedness over a finite set of alternatives allows us to conclude that there exists a preference relation that satisfies the core axioms. Clearly, the preferences of a subject over the whole space of payoffs need not be consistent with the core axioms even if his ranking is single-peaked over the finite set of alternatives used in the experiment. The extent of non-single-peakedness observed in our experiments will thus provide a *conservative* estimate.

Indifference curves may have linear segments if preferences are weakly convex. We elicit strict rankings in our experiment. Subjects may be indifferent between some of the alternatives and, when we compel them to report a strict ranking, the manner in which subjects break ties could lead them to report a non-single-peaked strict ranking. In particular, subjects who treat own payoff and other's payoff as perfect substitutes can report any ranking over a finite number of divisions of a fixed amount in the dictator game.

From a practical perspective, we elicit strict rankings primarily for convenience. Our elicitation method can be easily modified whereby subjects can be asked to report indifferences. From an empirical perspective, Andreoni and Miller (2002) and Fisman, Kariv, and Markovits (2007) report that the fraction of subjects who treat own and other's payoffs as perfect substitutes in standard dictator games is 6.2% and 2.63%, respectively. Finally, from a comparative-methodological perspective, the possibility of indifference is as much a concern when *choices* are elicited in one-shot interactions. If we entertain the possibility of indifference, then we can not conclude that a subject who gives half of the pie to his recipient is more altruistic than one who gives nothing in a one-shot dictator game where choices are elicited. We therefore believe that eliciting strict rankings only marginally affects the generality of our study.

3. Experimental Design and Procedures

We employ five treatments (Table 1) that are variants of dictator and trust games (). The experiments were conducted at the laboratory of the Max Planck Institute of Economics, Jena, Germany, from March 2011 to June 2011. Subjects were students at the University of Jena. We used a between-subjects design and the games were played only once in each session. We have data on 96 subjects (48 men and 48 women) for each treatment.

3.1. Treatments

Each alternative in treatment DB (Dictator-Baseline) prescribes a division of 20 between the dictator and the recipient. The dictator has to strictly rank the five alternatives. Note that any n alternatives can be strictly ranked in $n!$ ways where $n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$. Out of these $n!$ rankings only 2^{n-1} rankings will be single-peaked (Craven, 1992).

Table 1: The alternatives to be ranked by dictators and the trustees

Treatment	The five alternatives				
	a^1	a^2	a^3	a^4	a^5
1. DB - Dictator-Baseline	(0, 20)	(2, 18)	(4, 16)	(7, 13)	(9, 11)
2. TB - Trust-Baseline	(0, 20)	(2, 18)	(4, 16)	(7, 13)	(9, 11)
3. TE - Trust-Equality	(0, 20)	(2, 18)	(4, 16)	(7, 13)	(10, 10)
4. DM - Dictator-Monotonicity	(0, 20)	(4, 16)	(7, 13)	(10, 10)	(16, 11)
5. TM - Trust-Monotonicity	(0, 20)	(4, 16)	(7, 13)	(10, 10)	(16, 11)

Notes: The first entry in each alternative is the payoff of the recipient (trustor) in treatments based on the dictator (trust) game. The payoffs refer to actual euro amounts used in the experiments.

The five alternatives in DB lie on a downward sloping straight line in the two-dimensional space of monetary payoffs to *self* and *other*. However, they can be ordered in terms of the payoff to either self or other on a one-dimensional axis. Consequently, only 16 out of the 120 possible strict rankings (∼13%) in DB will be single-peaked. It is only these 16 single-peaked rankings that are permitted by the theoretical models of outcome-based other-regarding preferences. Intention-based theories have no bite in the dictator game.

Table 2 lists all the single-peaked rankings in treatment DB. We represent each alternative in terms of the amount that can be given by the dictator to the recipient. A subject who reports the ranking $r_1 \equiv 0 \succ 2 \succ 4 \succ 7 \succ 9$ is the so called 'neoclassical' subject and will be labeled as self-regarding. A subject who reports any of the remaining fifteen single-peaked rankings can be thought of as exhibiting other-regarding preferences in line with the existing theories.

Treatment TB (Trust-Baseline) was conducted to check the robustness of the results obtained from treatment DB. The game played by subjects in treatment TB is a variant of the standard trust game. The first mover – the trustor – has to choose either OUT or IN. If he chooses OUT, then the game ends with the trustor receiving 5 and the trustee receiving 0. If the trustor chooses IN, then the trustee is asked to rank the five possible alternatives which are identical to the five alternatives faced by the dictator in DB. The task faced by a trustee in TB is identical to that faced by a dictator in DB. Thus, 16 out of the 120 possible strict rankings of the five alternatives in TB will be single-peaked.

Table 2: Single-peaked rankings in treatment DB

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}
g_1	0	2	2	2	2	4	4	4	4	4	4	7	7	7	7	9
g_2	2	0	4	4	4	2	2	2	7	7	7	4	4	4	9	7
g_3	4	4	0	7	7	0	7	7	2	2	9	2	2	9	4	4
g_4	7	7	7	0	9	7	0	9	0	9	2	0	9	2	2	2
g_5	9	9	9	9	0	9	9	0	9	0	0	9	0	0	0	0

Notes: These rankings are the 16 single-peaked rankings in treatment TB as well. g_i refers to the amount given by self to other according to the i^{th}-ranked alternative. Replacing '9' by '10' throughout the table gives the list of the 16 single-peaked rankings in treatment TE.

Our objective is to examine whether there is any difference in the extent of single-peakedness across DB and TB. No existing theory of outcome-based or intention based other-regarding preferences would predict a difference in the extent of single-peakedness across the rankings submitted by dictators in DB and trustees in TB.[5] Differences in first-ranked alternatives (which may be thought of as a proxy for the choice subjects would have made if asked to choose one of the five alternatives) or in the distribution of rankings across these two treatments are of little importance for us given the focus of our analysis.

Treatment TE (Trust-Equality) differs from treatment TB in only one respect: alternative $(9, 11)$ in TB has been replaced with $(10, 10)$ in TE. As in treatments DB and TB, 16 out of the 120 possible strict rankings of the five alternatives in TE will be single-peaked. Prior research has shown that minor changes to the set of alternatives can have significant impact on choices made by subjects (Güth, Huck, and Müller, 2001; List, 2007, Bardsley, 2008). We shall argue later that such results can potentially be accommodated in the framework of Cox et al. (2008). Our interest lies in examining whether there is any difference in the extent of single-peakedness across TB and TE. No existing theory of other-regarding preferences would predict such a difference.

The reason for using treatments DM and TM will be explained shortly.[6] For the moment, note that payoffs to *self* and *other* sum to 20 in the first four alternatives but not in the fifth alternative which involves giving 16 to *other* and keeping 11 for *self*.

[5]This claim is obvious for outcome-based theories. In case of intention based theories, the reader may verify that the specification of the utility function is such that preferences over alternatives will be single-peaked for any fixed second-order belief of the trustee.

[6]Subjects can be classified in a more refined manner in DM and TM because we can identify non-monotonicity of preferences. This, in turn, permits a better analysis of *who* violates single-peakedness. The choice of these treatments was inspired by the treatment involving step-shaped budget sets in Fisman, Kariv, and Markovits (2007).

The five alternatives in treatments DM and TM can also be strictly ranked in 120 ways. However, only the first four alternatives whose payoffs sum to 20 lie on a straight line.[7] Only these four alternatives can be used to discuss the notion of single-peakedness. These four alternatives can be ordered on a one-dimensional axis in terms of the payoff to *self* (or *other*) and allow for 8 single-peaked rankings (Table 3).

Consider any one of these 8 single-peaked rankings, say, $r_7^m \equiv 7 \succ 10 \succ 4 \succ 0$. The fifth alternative – $(16, 11)$ – can potentially be placed at five locations to arrive at a strict ranking of all the five alternatives: it can be above giving 7, below giving 0, or anywhere in between. Each of the resulting five strict rankings will be consistent with the core axioms (see Figure 4). The crucial thing to note is that preferences will be non-monotonic with respect to the payoff of *other* if alternative $(10, 10)$ is ranked above $(16, 11)$ irrespective of the exact placement of these two alternatives in the overall ranking of the subject (see Figures 4(C), 4(D), and 4(E)).

In general, there exist five theoretically consistent strict rankings for each of the 8 single-peaked rankings of the four alternatives that sum to 20. Thus, 40 out of the 120 possible strict rankings of all the five alternatives in treatments DM and TM are consistent with the theoretical framework of other-regarding preferences (see Table S1 in the Supplement). There should be no difference in the fraction of rankings consistent with the existing theories when we compare dictators in DM with trustees in TM.

3.2. Implementation

Experiments were conducted in a networked computer laboratory. The experimental games were programmed using the software Zurich toolbox for ready-made economic experiments (z-Tree) developed by Urs Fischbacher (2007).

3.2.1. General procedures

During each session, while entering the lab, each subject picked a card indicating the computer terminal number. The lab can accommodate 32 subjects and so excess subjects were given the show-up fee of €2.50 and asked to leave. Instructions for the experiment were distributed after all the subjects had been seated. An experimenter read the instructions aloud after all subjects finished reading instructions on their own. Participants were then asked to answer a questionnaire to check their understanding of the instructions. The experiment started after all participants correctly answered all the questions.[8]

[7] The fact that the $(16, 11)$ alternative and any *one* of the first four alternatives will also lie on a line is inconsequential for our discussion.

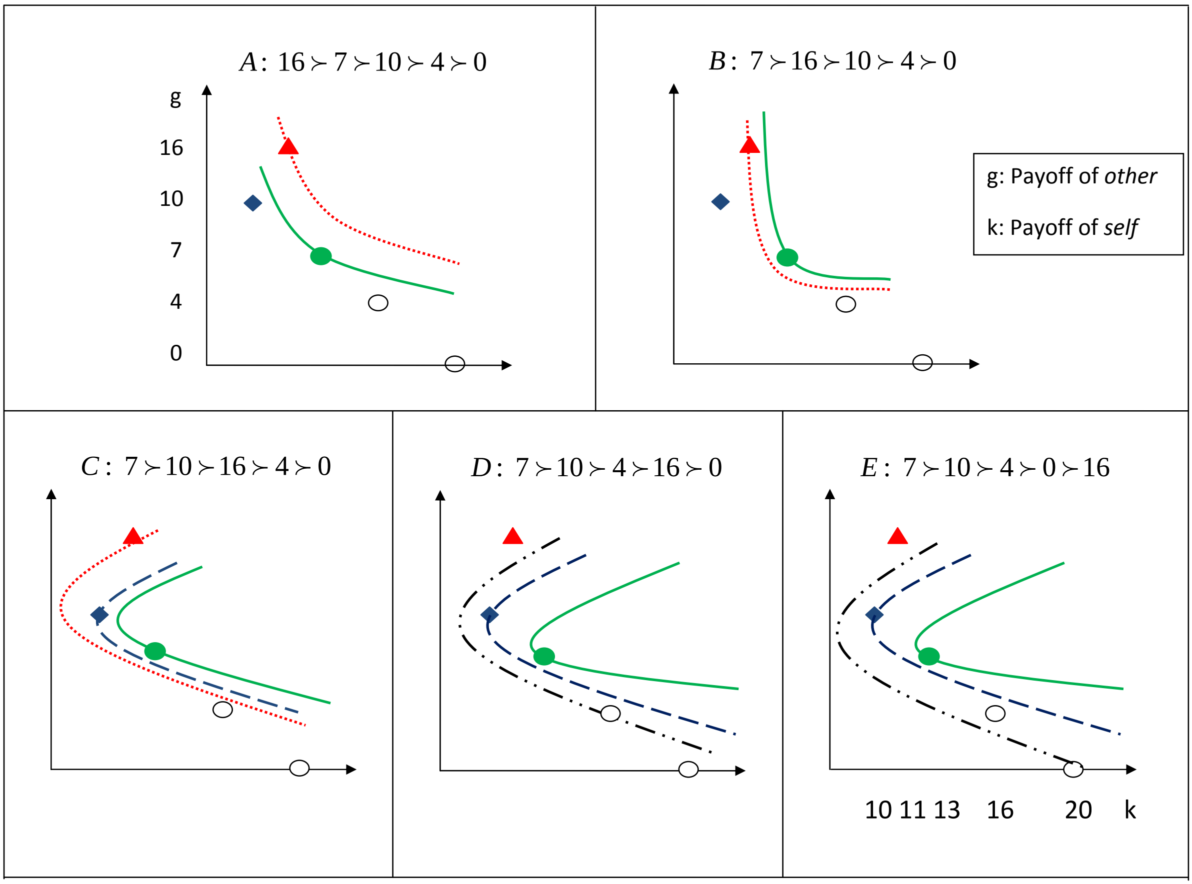
[8] The number of subjects who were not able to answer all the questions on the first attempt varied from 1 to 3 (out of 32) across all the sessions.

Table 3: Single-peaked rankings of the first four alternatives in treatments DM and TM

Rank	r_1^m	r_2^m	r_3^m	r_4^m	r_5^m	r_6^m	r_7^m	r_8^m
g_1^m	0	4	4	4	7	7	7	10
g_2^m	4	0	7	7	4	4	10	7
g_3^m	7	7	0	10	0	10	4	4
g_4^m	10	10	10	0	10	0	0	0

g_i^m refers to the amount given by self to other according to the i^{th}-ranked alternative among the first four alternatives whose payoffs sum to 20.

Figure 4: An example of the five theoretically feasible preference rankings of all the five alternatives that are consistent with the r_7^m ranking of the four alternatives other than the $(16, 11)$ alternative.



The remaining three treatments – TB, TE, and TM – are based on the trust game. The *strategy-vector* method was used in these treatments. Both subjects in each pair first ranked the alternatives in the role of the trustee with the knowledge of the probabilities of implementing the payoffs. Then both subjects made the choice between OUT and IN in the role of the trustor without any information about decisions made by other subjects.

One card was drawn from the box containing the 32 terminal IDs when all subjects finished making all the decisions in any given session. The participant sitting on the terminal corresponding to drawn card was asked to come forward. (S)he was then shown a box containing 100 cards (50 red, 30 blue, 15 black, and 5 white). The box was closed, vigorously shaken, and the participant was asked to draw one card from the box. The color of the card drawn by the participant determined which rank will be used to determine the payoffs for all participants in the session. For example, if a blue card was drawn in a session, then the second-ranked alternative was used to determine the payoffs for all pairs.

The *actual* roles (dictator/recipeint and trustor/trustee) were then randomly determined by the computer. Both subjects in a given pair were paid according to the ranking submitted by the *actual* dictator in treatments based on the dictator game. In treatments based on the trust game, the ranking provided by the actual trustee was used to determine payoffs if the actual trustor chose IN. If the actual trustor had chosen OUT, then he got 5 and the actual trustee got 0. Payoffs were displayed on the computer screen of each subject. Subjects were then asked to approach the experimenter one by one in the order of their terminal IDs, collect their earnings in cash, and leave the lab.

Before describing the details of the ranking elicitation mechanism, we discuss three potential concerns regarding the experimental design. First, it is well known that choices made by trustees may depend upon whether one uses the 'play' method or the 'strategy' method (Casari and Cason, 2010). The strategy method and the play method may lead to differences in first-ranked alternatives and/or the distribution of rankings but there is no theoretical reason to believe that single-peakedness of rankings would be affected. None of the existing theories of other-regarding preferences predict such a difference. Hence, there does not seem to be any reason not to use the strategy method.

Second, the actual roles of subjects were determined after both subjects in a pair made decisions as dictators (trustees and trustors) in treatments based on the dictator (trust) game. Whatever concerns could arise from this design feature are minor issues from a theoretical perspective as they can be easily eliminated by conducting experiments where roles are assigned apriori.[9] Finally, our experiment was single-blind. Once again, there is no compelling theoretical reason to believe that single-peakedness would be sensitive to this design feature.

[9]Subjects were paid according to their decision in one of the two roles and thus 'portfolio effects' should not be a concern (Cox, 2010).

3.2.2. Ranking-elicitation mechanism

In each treatment subjects were told that both subjects in the pair will be paid according to the first, second, third, fourth, or fifth ranked alternative of the actual dictator (or the actual trustee) with *monotonically decreasing probabilities* of 0.50, 0.30, 0.15, 0.05, and 0.00. We shall, henceforth, refer to this mechanism as the MDP mechanism.[10] Subjects report one out of 120 possible rankings of the five alternatives in every treatment. Reporting a ranking effectively amounts to choosing one one out of 120 possible lotteries where each possible outcome of the lottery specifies monetary payoffs for *self* and *other*. Since the games are played only once, providing a ranking is equivalent to *one-task* lottery choice experiments. Cox, sadiraj and Schmidt (2011) clarify that, in theory, the one-task mechanism is the best available way to induce subjects to report their preferences over lotteries truthfully since it is immune to all concerns (wealth effects and portfolio effects) that arise in *multi-task* lottery choice experiments. Hence, we find it reasonable to work under the assumption that the MDP mechanism will provide the incentives for truthfully revealing preferences over lotteries, irrespective of the structure of those preferences. The key features of the MDP mechanism are best understood by answering the following two questions.

[**Q1**] What should be the structure of preferences over lotteries so that the reported ranking of alternatives can be unambiguously interpreted as the preference ranking over the sure alternatives?
[**Q2**] What must be the structure of the inferred ranking over sure alternatives so that it is in line with the existing theories of other-regarding preferences?

To answer [Q1], suppose that a subject behaves in accordance with *some* theory of decision making under risk according to which lotteries are evaluated using special cases of the functional

$$V = \sum_{i=1}^{5} \pi_i(\mathbf{p}) \cdot u(a_i) = \sum_{i=1}^{5} \pi_i(\mathbf{p}) \cdot u(g_i, k_i), \tag{1}$$

where
(i) $\pi_i(\mathbf{p})$ is the decision weight on the i^{th}-ranked alternative which may depend on the whole vector of the *objectively* given probabilities $\mathbf{p} = (p_1, p_2, p_3, p_4, p_5)$; and,
(ii) $u(a_i) = u(g_i, k_i)$ is the utility derived from the i^{th}-ranked alternative a_i which prescribes giving g_i to other and keeping k_i for self.

[10] We thank Tim Cason for drawing our attention to the work of Chakravarty, Ma, and Maximiano (2011) who have independently used this mechanism to study a different question.

Note that subjects behaving in line several theories of decision making under risk would evaluate lotteries according to the above functional (for example, Expected Utility Theory (Morgenstern and Von Neumann, 1944), Prospect Theory (Kahnemann and Tversky, 1979), Rank Dependent Expected Utility Theory (Quiggin, 1982), and the dual theory of Expected Utility Theory (Yaari, 1987)).

Suppose lotteries are evaluated according to the above mentioned functional. The reported ranking of alternatives can be unambiguously interpreted as the preference ranking over the sure alternatives if the decision weights are monotonically decreasing, i.e., if $\pi_i(\mathbf{p}) > \pi_j(\mathbf{p})$ for $1 \leq i < j \leq 5$. This requirement clearly holds for subjects who behave according to Expected Utility Theory *irrespective* of their attitudes towards risk since $\pi_i(\mathbf{p}) = p_i$, for all $1 \leq i \leq 5$, in such a case.

If a subject in our experiment behaves in accordance with *any* theory of decision making under risk wherein (i) lotteries are evaluated using the above mentioned functional and (ii) the decision wights are monotonically decreasing, then the answer to [Q2] is obvious: the inferred ranking over sure alternatives must be single-peaked for it to be consistent with existing theories of other-regarding preferences.

4. Hypotheses

4.1. Extent of single-peakedness

The distribution of first-ranked alternatives may differ across treatments. The distribution of rankings may also differ. However, if the analytical framework of other-regarding preferences is empirically sound, no subject should report a non-single-peaked ranking in any treatment. Even if we acknowledge the possibility of random errors, there should be no difference in the extent of single-peakedness across (i) DB and TB, (ii) TB and TE, and (iii) DM and TM. This is because the number of single-peaked rankings for both treatments within each of these three pairwise comparisons is identical.

One of our major interests lies in comparing treatment TB which contains the $(9, 11)$ alternative with treatment TB which instead contains $(10, 10)$. We are *not* interested in examining whether there is a difference in the distribution of first-ranked alternatives across TB and TE. Differences in first-ranked alternatives can potentially be accommodated by allowing not only the choices of others but also the structure of the set of alternatives to affect whether an agent adopts a more or less altruistic preference structure in accordance with the framework of Cox et al. (2008).

In the terminology of Cox et al. (2008), the mere presence of $(10, 10)$ in place of $(9, 11)$ in the set of alternatives may lead the trustees to adopt a relatively 'more (less) altruistic' preference structure. Thus, choices (or, first-ranked alternatives) could differ across these two treatments. Nonetheless, the preferences under any scenario will be single-peaked if the core preference axioms are satisfied. Hence, we shall not consider a significant difference in the distribution of first-ranked alternatives across TB and TE as being outside the scope of the existing framework. Our test for the empirical validity of the framework will be that there should be no difference in the extent of single-peakedness across TB and TE.

We would also like to add that if the existing framework captures the preference structures of men and women equally well, then the distribution of first-ranked alternatives and/or rankings of men and women may be different in any given treatment but the extent of single-peakedness should not be different.

4.2. Who violates single-peakedness

We shall follow the terminology introduced by Fehr and Schmidt (1999) and classify subjects in terms of their aversion towards advantageous and disadvantageous inequality. Our results do not depend in any way on this choice of the classification scheme. We use it because it provides a simple language to classify subjects across all our treatments.

First, consider treatments DB, TB, and TE where each alternative sums to 20 and leads to a situation of (strictly or weakly) *advantageous inequality* for the decision maker. Let g_1 denote the amount given by a subject to other according to his first-ranked alternative in treatments DB, TB, and TE. Subject S_1 will be said to exhibit a stronger aversion to advantageous inequality (henceforth, AAI) than subject S_2 if $g_1(S_1) > g_1(S_2)$. Note that knowing only the first-ranked alternative does not allow us to infer whether the overall ranking reported by a subject is single-peaked or not.

It is important to note that the way we classify subjects is in line with how subjects would be classified if they were asked to choose one of the five alternatives rather than rank them. Our analysis of who violates single-peakedness will be based on the assumption that the first-ranked alternative in our treatments would be the choice subjects would make if they were asked to choose one of the five alternatives.[11]

Let g_1^m denote the amount given by a subject to other according to his first-ranked alternative among the four alternatives that sum to 20 in treatments DM and TM. In treatments DM and TM, the first four alternatives lead to a situation of (strictly or weakly) advantageous inequality for the decision maker whereas the fifth alternative –$(16, 11)$ – leads to a situation of disadvantageous inequality.

[11] We believe this is a reasonable assumption. Nonetheless, we conducted two additional treatments (using 64 subjects) that were similar to treatments DB and TM except that subjects had to choose an alternative. We found no difference in the distribution of first-ranked alternatives in the ranking version of treatment DB (TM) and the distribution of choices in its choice version [$P = 0.948$, two-sided Mann-Whitney U test] ([$P = 0.817$, two-sided Mann-Whitney U test]).

We can classify subjects in terms of AAI on the basis of g_1^m in treatments DM and TM. In addition, we can also classify subjects in these two treatments in terms of their aversion to disadvantageous inequality (henceforth, ADI). Note that preferences are non-monotonic with respect to the payoff of *other* if alternative $(10, 10)$ is ranked above $(16, 11)$, irrespective of the exact placement of these two alternatives in the overall ranking. Revealing non-monotonicity with respect to the payoff of *other* is observationally equivalent to revealing ADI. In contrast, if a subject ranks alternative $(16, 11)$ above $(10, 10)$, then he has no (or very low) ADI.

As an example, suppose the rankings of subjects S_3 and S_4 in treatment DM in terms of the amount given to *other* turn out to be $7 \succ 4 \succ 10 \succ 0 \succ 16$ and $16 \succ 4 \succ 10 \succ 7 \succ 0$, respectively. S_3 will be classified as being relatively more AAI than S_4 because $g_1^m(S_3) = 7 > 4 = g_1^m(S_4)$. In addition, subject S_3 shows ADI since he ranks $(10, 10)$ above $(16, 11)$, whereas subject S_4 does not.[12]

To summarize, we shall classify subjects into a total of five categories based on their AAI in treatments DB, TB, and TE. In treatments DM and TM, subjects will be classified into eight categories depending on the four possible levels of AAI and two levels of ADI. Formally, we can define the variables AAI and ADI as follows.

$$\text{AAI} = \begin{cases} g_1 & \text{in treatments DB, TB, and TE} \\ g_1^m & \text{in treatments DM and TM} \end{cases} \tag{2}$$

$$\text{ADI} = \begin{cases} 0 & \text{if } (16,11) \succ (10,10) \text{ in treatments DM and TM} \\ 1 & \text{if } (10,10) \succ (16,11) \text{ in treatments DM and TM} \end{cases} \tag{3}$$

Our final hypothesis is that there should be no difference in the extent of single-peakedness across subsets of subjects that are classified differently in terms of AAI in treatments DB, TB, and TE; and, in terms of both AAI and ADI in treatments DM and TM. In particular, subjects who show AAI, or ADI, or both, should be as likely to exhibit single-peakedness as those who will be 'labeled' as 'neoclassical' subjects. We stress the word 'labeled' because, in terms of the amount given to *other*, a truly neoclassical subject should report (i) the ranking $0 \succ 2 \succ 4 \succ 7 \succ 9$ in treatments DB and TB, (ii) the ranking $0 \succ 2 \succ 4 \succ 7 \succ 10$ in treatment TE, and (iii) the ranking $0 \succ 4 \succ 7 \succ 16 \succ 10$ in treatments DM and TM.

[12] The ranking of subject (S_4) S_3 is (not) permitted by existing theories because the ranking over the four alternatives other than $(16, 11)$ is (not) single-peaked. Of course, we could not arrive at this conclusion if we only knew g_1^m and the relative ranking of alternatives $(16, 11)$ and $(10, 10)$.

We classify subjects as they would be in experiments where choices are elicited and examine whether their overall ranking is single-peaked. We shall therefore label a subject as neoclassical if (i) he does not show AAI in treatments DB, TB, and TE and (ii) shows neither AAI nor ADI in treatments DM and TM. Put differently, a subject will be labeled as neoclassical if (i) $g_1 = 0$ in treatments DB, TB, and TE and (ii) $g_1^m = 0$ and alternative $(16, 11)$ is ranked above $(10, 10)$ in treatments DM and TM.

5. Data Analysis

We used a between-subjects design. Each session involved 32 different subjects. Three sessions of each treatment were conducted. Our data comprise of 96 observations for each treatment. The treatments were gender balanced and thus we have data on 48 men and 48 women in each treatment.[13] We shall treat each observation as an independent observation since the games were played only once during a session.

5.1. Overall extent of single-peakedness

Pooling data from all sessions we find that ∼75% of subjects (358 out of 480) report single-peaked rankings (Table 4). In fact, the fraction of single-peaked rankings is at least 79% in all treatments except the TE treatment. If subjects were to behave randomly, then the extent of single-peakedness would be ∼13% in treatments DB, TB, and TE, and ∼33% in treatments DM and TM. The hypothesis that subjects behave randomly is emphatically rejected for every treatment using a test of proportions. In addition, there is no statistical difference in the extent of single-peakedness across men and women in any treatment and in the pooled data. The fraction of men and women reporting single-peaked rankings is almost identical in every treatment and exactly identical in the pooled data.

5.2. Single-peakedness across treatments

The extent of single-peakedness in itself is perhaps not very informative because it could vary with design features like the cardinality of the set of alternatives. Treatment comparisons are thus necessary. As discussed earlier, there should be no difference in the extent of single-peakedness across (i) DB and TB, (ii) TB and TE, and (iii) DM and TM. Note that we are comparing only those treatments that permit equal number of theoretically feasible rankings.

[13]The sessions were not gender balanced. No restriction was placed on the gender composition while recruiting subjects for the first (or the first and second) session of any treatment. Restrictions were then imposed for the latter session(s) to ensure gender balance.

Table 4: Extent of single-peaked rankings across treatments

Data	Treatment					
	All	DB	TB	TE	DM	TM
N	$\frac{358}{480}$ (75%)	$\frac{76}{96}$ (79%)	$\frac{76}{96}$ (79%)	$\frac{52}{96}$ (54%)	$\frac{76}{96}$ (79%)	$\frac{78}{96}$ (81%)
M	$\frac{179}{240}$ (75%)	$\frac{39}{48}$ (81%)	$\frac{40}{48}$ (83%)	$\frac{25}{48}$ (52%)	$\frac{37}{48}$ (77%)	$\frac{38}{48}$ (79%)
W	$\frac{179}{240}$ (75%)	$\frac{37}{48}$ (77%)	$\frac{36}{48}$ (75%)	$\frac{27}{48}$ (56%)	$\frac{39}{48}$ (81%)	$\frac{40}{48}$ (83%)

Notes: N, M, and W refer to data on all subjects, data only on men, and data only on women, respectively. The fraction reported in the table is ratio of number of subjects who report a single-peaked ranking and the total number of subjects. The numbers inside the parentheses report the corresponding percentage rounded off to the nearest integer.

There is no statistical difference in the extent of single-peakedness across treatments DB and TB [$\chi^2(1) = 0.000$, $P = 1.000$]. The same result holds for the comparison between treatments DM and TM [$\chi^2(1) = 0.131$, $P = 0.717$].[14] Both these results hold if we look at the data on men and women separately. The extent of single-peakedness is, however, significantly different across treatments TB and TE [$\chi^2(1) = 13.5$, $P < 0.001$].

The presence of $(10, 10)$ impacts the behavior of subjects in treatment TE in several ways. For instance, we find weak difference in the distribution of first-ranked alternatives across these treatments [$P = 0.062$, two-sided Mann-Whitney U test]. This finding is similar to the results of Güth, Huck, and Müller (2001) using the ultimatum game (Güth, Schmittberger, and Schwarze, 1982). However, such findings can potentially be accommodated in an extended version of the Cox et al. (2008). It is the significant difference in the extent of single-peakedness across TB and TE which is hard to reconcile.

Table 5 presents the rank-distribution of alternatives $(9, 11)$ and $(10, 10)$ in treatments TB and TE, respectively. Alternative $(10, 10)$ is ranked *fifth* in treatment TE with a significantly lower frequency than alternative $(9, 11)$ is in TB [$\chi^2(1) = 11.032$, $P = 0.001$]. This result holds separately for men [$\chi^2(1) = 4.174$, $P = 0.041$] and women [$\chi^2(1) = 7.045$, $P = 0.008$]. In contrast, alternative $(10, 10)$ is ranked *first* in treatment TE with a significantly higher frequency than the alternative $(9, 11)$ is in treatment TB [$\chi^2(1) = 4.761$, $P < 0.029$]. Statistically, this result is driven by women [$\chi^2(1) = 5.275$, $P = 0.022$] since we do not find such an effect for men [$\chi^2(1) = 0.549$, $P = 0.459$].

[14]The extent of single-peakedness is identical across treatments DB and DM. However, this should be regarded as a coincidence since, as explained earlier, these two treatments are not truly comparable because they allow for different numbers of theoretically feasible single-peaked rankings.

Table 5: Rank distribution of $(9,11)$ in TB and $(10,10)$ in TE

Rank	(9,11) in TB			(10,10) in TE			$\Delta^{\dagger}$		
	N	M	W	N	M	W	N	M	W
1^{st}	17	9	8	30	12	18	13	3	10
2^{nd}	12	6	6	13	8	5	1	2	-1
3^{rd}	3	2	1	6	4	2	3	2	0
4^{th}	6	3	3	12	6	6	6	3	3
5^{th}	58	28	30	35	18	17	-23	-10	-13
Total	96	48	48	96	48	48	0	0	0

Notes: Δ reports the difference between the frequency of $(10,10)$ and $(9,11)$ at a given rank.

To summarize, both men and women seem to rank $(10,10)$ in TE relatively higher than $(9,11)$ in TB; but, women seem to be more likely to place alternative $(10,10)$ at the top of their ranking. These subtle differences, however, do not seem to influence the overall extent of single-peakedness across men and women in TE. Almost identical numbers of men and women violate single-peakedness in treatment TE.

5.3. Who violates single-peakedness?

A subject facing the task of ranking the alternatives probably has to resolve several questions. Just deciding which alternative to rank first may involve several considerations depending on the nature of the set of alternatives: whether to give anything to *other* or not; if so, how much; should one be willing to accept a lower monetary payoff than the other individual or forsake efficiency for equality? Whether the ranking is single-peaked or not is likely to be *determined by* the number and the nature of considerations a subject undertakes. A regression specification where single-peakedness is the dependant variable is perhaps less subject to endogeneity concerns. Hence, our analysis of who violates single-peakedness will be based on Probit regressions where the *dependent* variable takes the value (0) 1 if the ranking reported by a subject is (non) single-peaked.

Subjects are classified only in terms of their AAI in treatments DB, TB, and TE, while they are classified using both AAI and ADI in treatments DM and TM. We will therefore analyze the two sets of treatments separately. Our focus shall be on examining whether the likelihood of reporting a single-peaked ranking varies systematically across subjects who differ in terms of their AAI and/or ADI. Successful empirical validation of the analytical framework of other-regarding preferences requires that subjects who reveal positive and/or negative regard for payoffs of *other* (which is captured by AAI and ADI, respectively) should be as likely to report single-peaked rankings as are the neoclassical subjects.

5.3.1. Treatments DB, TB, and TE

Table 6 reports the marginal effects and robust standard errors from Probit regressions. The main result emerging from Model 1 is that, on average, the likelihood of single-peakedness is relatively lower for subjects who show higher AAI. Recall that the variable AAI stands for the amount given to *other* according to one's first-ranked alternative in treatments DB, TB, and TE. Thus, subjects who are relatively more generous according to their first-ranked alternative are relatively more likely to violate single-peakedness.

Model 1 confirms that the likelihood of single-peakedness does not differ across men and women. It also confirms that subjects in treatment TE are significantly less likely to be single-peaked relative to TB. We have already reported in Section 5.2 that subjects behave slightly more generously in terms of their first-ranked alternatives in treatment TE which contains the $(10, 10)$ alternative than in treatment TB which instead contains the $(9, 11)$ alternative. Does this feature drive the significantly lower incidence of single-peakedness in treatment TE relative to TB? Model 2 in Table 6 tests this hypothesis by including the interaction between the AAI variable and the TE treatment dummy. The statistical insignificance of the interaction term effectively rules out this hypothesis.

Table 7 reports the (average) predicted probability of being single-peaked for subjects differing in terms of AAI based on Model 2.[15] The predicted probability of single-peakedness is nearly 87% for the neoclassical subjects in DB or TB who gave 0 according to their first-ranked alternative. The corresponding probability for subjects who gave 9 according to their first-ranked alternative in DB or TB is 0.66. The probability of single-peakedness is uniformly lower in treatment TE relative to treatments TB and DB.[16] It ranges from a high of 0.64 for the neoclassical subjects to a low of 0.40 for subjects who gave 10 according to their first-ranked alternative.

Models 1 and 2 produce consistent results: subjects who reveal a relatively stronger AAI by giving more to *other* according to their first-ranked alternative are, on average, relatively less likely to be single-peaked. This suggests that *generosity runs out* more often than *selfishness runs out*.

Is this result a behavioral regularity or merely a statistical one? Consider treatment DB where a subject can give 0, 2, 4, 7, or 9 to *other* according to his first-ranked alternative.

[15]Predicted probability of being single-peaked is calculated for each individual using the estimated coefficients. Subjects with the same AAI are pooled together and the average of their predicted probabilities is reported.

[16]We discuss our preferred explanation for this difference in the concluding section.

Table 6: Who violates single-peakedness in DB, TB, and TE?

	Model 1	Model 2
DB	0.027 (0.068)	0.031 (0.068)
TE	-0.228*** (0.071)	-0.255*** (0.091)
Male	0.016 (0.054)	0.017 (0.054)
AAI	-0.023*** (0.007)	-0.026*** (0.009)
TE × AAI		0.006 (0.013)
Observations	288	288
Pseudo-R^2	0.0896	0.0890

Notes: The table reports marginal effects of probit estimations. The dependent variable takes the value (0) 1 if an individual's ranking is (non) single-peaked. DB and TE are treatment dummies. AAI refers to aversion towards advantageous inequality and takes five possible values. Treatment TB is the baseline. Robust standard errors are in parentheses. Constant included. * – P < 0.1; ** – P < 0.05; *** – P < 0.01.

Table 7: Predicted probability of single-peakedness based on Model 2

Treatment	Increasing AAI →				
	$g_1 = 0$	$g_1 = 2$	$g_1 = 4$	$g_1 = 7$	$g_1 = 9, 10$
DB, TB	0.87 (36, 47)	0.83 (10, 10)	0.79 (9, 9)	0.72 (17, 13)	0.66 (24, 17)
TE	0.64 (45)	0.59 (6)	0.54 (5)	0.48 (10)	0.40 (30)

Notes: g_1 refers to the amount given to other according to the first-ranked alternative. Treatment TE contains the $(10, 10)$ alternative instead of the $(9, 11)$ alternative. The reported probabilities are based on Model 2 presented in Table 7. The numbers in parentheses report the number of subjects in the corresponding cell of the table. The qualitative results do not change if we pool subjects in the three middle categories of AAI where $g_1 \in \{2, 4, 7\}$.

Irrespective of which alternative is ranked first, the remaining four alternatives can be arranged in 24 ways. Now suppose the first-ranked alternative is giving 0 to other. Out of the 24 possible arrangements of the remaining four alternatives only 1 will make the overall ranking of the five alternatives single-peaked. In general, 1, 4, 6, 4, and 1 out of the 24 arrangements of the remaining four alternatives will make the overall ranking single-peaked when the first-ranked alternative is giving 0, 2, 4, 7, or 9, respectively (see Table 2). Clearly, subjects who give relatively more to *other* according to their first-ranked alternative are not *uniformly* less likely to exhibit single-peakedness based on purely statistical considerations. We, therefore, believe that this result reveals a behavioral, and not, a statistical regularity.

The regression results do not lend support to another intuitively appealing view regarding the likelihood of single-peakedness. It is tempting to think that subjects who start with giving 0 or giving 9 (or, 10 in TE) to *other* as their first-ranked alternative would be more likely to be single-peaked because of the relative ease of coming up with a ranking. This is true for subjects whose first ranked alternative is giving 0. However, subjects who start with giving 9 (or, 10) to *other* as their first-ranked alternative are less likely to be single-peaked than those who start with giving 2, 4, or 7. Thus, the ease of coming up with a ranking does not provide a coherent explanation for the patterns observed in the data.

5.3.2. Treatments DM and TM

The analysis of treatments DB, TB, and TE revealed that *positively* other-regarding tendencies, as captured by aversion to advantageous inequality, reduce the likelihood of single-peakedness. Treatments DM and TM allow us to examine whether the same holds with respect to *negatively* other-regarding tendencies which are captured by aversion to disadvantageous inequality.

Preferences of a subject will be non-monotonic with respect to the payoff of *other* if alternative $(10, 10)$ is ranked above $(16, 11)$ irrespective of the exact placement of these two alternatives in the overall ranking. A non-monotonic preference ranking is equivalent to revealing ADI. As discussed earlier, treatments DM and TM allow a more refined classification of subjects. Subjects can be classified both in terms of their AAI and ADI. Before analyzing the Probit regressions we note that about 30% of the subjects rank the alternative $(10, 10)$ above $(16, 11)$ in treatments DM and TM and thus exhibit non-monotonicity (see Table A4 in the Supplement). There is no difference in the proportion of subjects who exhibit non-monotonicity across treatments DM and TM [$\chi^2(1) = 0.099$, $P = 0.753$].[17]

[17]The proportion of women who exhibit non-monotonicity (∼40%) is higher than the corresponding proportion of men (∼20%). This result is in line with those reported by Andreoni and Vesterlund (2002) and surveyed by Croson and Gneezy (2009).

Results of Probit regressions for treatments DM and TM are reported in Table 8. Model 1 reveals significantly negative marginal effects of AAI and ADI. The results from Model 1 can be summarized as follows.

1. Increasing AAI reduces the likelihood of single-peakedness irrespective of whether subjects show ADI or not.

2. Subjects who show ADI are less likely to be single-peaked than those who do not, at *all* levels of AAI.

Consider one of the key differences between treatments DB and DM. A subject in treatment DM faces the additional consideration of how to rank the efficient $(16, 11)$ alternative relative to the inefficient but 'equal' $(10, 10)$ alternative; a consideration which is absent in treatment DB given the set of alternatives. The manner in which a subject resolves this consideration determines whether or not we label him as ADI in treatment DM. It is possible that the marginal impact of this consideration on the likelihood of single-peakedness varies with the level of AAI. For instance, the marginal effect could be relatively lower for subjects who reveal stronger AAI as such subjects are already quite likely to be non-single-peaked based on the findings from treatments DB. There is no theory to suggest that the opposite case is not possible. Hence, Model 2^m reported in Table 8 examines this issue by including the interaction of AAI and ADI as an additional regressor.

The positive and statistically significant marginal effect of the interaction term in Model 2^m helps us revise the findings from Model 1^m in the following manner (see Table 9 for the predicted probabilities of single-peakedness based on the results from Model 2^m).

1. Increasing AAI reduces the likelihood of single-peakedness among subjects who do not show ADI but has no impact on those who show ADI.

2. Subjects who show ADI are less likely to be single-peaked than those who do not at *all but the highest* level of AAI (which is confirmed by t-tests at the 5% significance level).

The common result from both Models 1^m and 2^m is that that having a high AAI *or* being ADI decreases the likelihood of single-peakedness. In other words, presence of positive and/or negative other regarding tendency reduces the likelihood of single-peakedness relative to the absence of both which goes against the implicit assumption of the existing theories of other-regarding preferences.

Table 8: Who violates single-peakedness in DM and TM?

	Model 1^m	Model 2^m
TM	0.067	0.068
	(0.057)	(0.057)
Male	-0.086	-0.073
	(0.056)	(0.056)
AAI	-0.023***	-0.032***
	(0.007)	(0.008)
ADI	-0.156**	-0.328***
	(0.071)	(0.116)
AAI × ADI		0.026*
		(0.013)
Observations	192	192
Pseudo-R^2	0.0961	0.1160

Notes: Same as Table 8 with the appropriate modifications in view of the treatments involved.

Table 9: Predicted probability of single-peakedness based on Model 2^m

	Increasing AAI →			
	$g_1^m = 0$	$g_1^m = 4$	$g_1^m = 7$	$g_1^m = 10$
Not ADI	0.95	0.89	0.78	0.66
	(61)	(19)	(20)	(34)
ADI	0.72	0.70	0.72	0.69
	(18)	(16)	(4)	(20)

Notes: g_1^m refers to the amount given to other according to the first-ranked alternative among the four alternatives excluding $(16, 11)$. A subject is (not) ADI if he ranks alternative $(16, 11)$ (above) below alternative $(16, 11)$. The numbers in parentheses report the number of subjects in the corresponding cell of the table. The qualitative results do not change if we pool subjects in the two middle categories of AAI where $g_1^m \in \{4, 7\}$.

Table 10: Single-peakedness across neoclassical and non-neoclassical subjects

Type of Subjects	Treatment					
	All	DB	TB	TE	DM	TM
NC	$\frac{159}{189}$ (84%)	$\frac{33}{36}$ (92%)	$\frac{41}{47}$ (87%)	$\frac{28}{45}$ (62%)	$\frac{36}{39}$ (92%)	$\frac{21}{22}$ (95%)
non-NC	$\frac{199}{291}$ (68%)	$\frac{43}{60}$ (72%)	$\frac{35}{49}$ (71%)	$\frac{24}{51}$ (47%)	$\frac{40}{57}$ (70%)	$\frac{57}{74}$ (77%)

Notes: (non) NC stands for (non) neoclassical subjects. The entries in the table should be read as follows: 159 out of the 189 subjects across all treatments who can be classified as NC subjects reported a single-peaked ranking. The numbers in the parentheses report the corresponding percentage rounded off to the nearest integer.

We conclude the discussion of who violates single-peakedness by reporting the proportion of neoclassical and non-neoclassical subjects who exhibit single-peakedness (Table 10). We have labeled subjects as neoclassical if (i) $g_1 = 0$ in treatments DB, TB, and TE and (ii) $g_1^m = 0$ and $(16, 11)$ is ranked above $(10, 10)$ in treatments DM and TM. All remaining subjects belong to the non-NC category. About 84% of neoclassical subjects and 68% of non-neoclassical subjects exhibit non-single-peakedness in the pooled data.

5.4. Where does non-single-peakedness arise?

We wrap up the data analysis by highlighting another interesting pattern in the data. Consider treatments DB, TB, and TE where the notion of single-peakedness is defined over all the five alternatives. We shall refer to an alternative in terms of the amount that can be given to *other*. Suppose a subject's first-ranked alternative is, say, 4. If his second-ranked alternative is either 0 or 9, then even without knowing the rest of his ranking we can unambiguously conclude that his overall ranking will be non-single-peaked. In such a scenario, we shall say that the subject exhibited non-single-peakedness *for the first time* at his 'second-ranked' alternative. In contrast, if his second-ranked alternative is 2 or 7 we can not yet conclude that his overall ranking is single-peaked or non-single-peaked.

Suppose the subject's first two alternatives are $4 \succ 2$. If the third-ranked alternative is 9, then the subject's ranking will definitely be non-single-peaked and we shall say that the subject exhibited non-single-peakedness for the first time at his 'third-ranked' alternative. In contrast, if his third-ranked alternative is 7 or 0, then we can not yet be conclusive. Now suppose the first three alternatives are $4 \succ 2 \succ 0$. If the fourth-ranked alternative is 9, then his ranking will be non-single-peaked and we shall say that he exhibited non-single-

Table 11: Where does non-single-peakedness occur for the first time?

First violation of SPness occurs at	Treatments			
	DB	TB	TE	Total
2^{nd} rank	$\frac{10}{20}$ (50%)	$\frac{14}{20}$ (67%)	$\frac{25}{44}$ (57%)	$\frac{49}{84}$ (58%)
3^{rd} rank	$\frac{8}{20}$ (40%)	$\frac{6}{20}$ (33%)	$\frac{10}{44}$ (23%)	$\frac{24}{84}$ (29%)
4^{th} rank	$\frac{2}{20}$ (10%)	$\frac{0}{20}$ (0%)	$\frac{9}{44}$ (20%)	$\frac{11}{84}$ (13%)

Notes: The entries in the table should be read as follows: 10 out of the total 20 violations of single-peakedness in treatment DB occurred at the second rank. The numbers in the parentheses report the corresponding percentage rounded off to the nearest integer.

peakedness for the first time at his 'fourth-ranked' alternative. In contrast, if his first four alternatives are $4 \succ 2 \succ 0 \succ 7$, then his ranking must be single-peaked.

In general, in treatments DB, TB, and TE, we can go through the subject's ranking and identify where he exhibited non-single-peakedness for the first time. Any subject can do so at the second, third, or fourth rank. Table 11 reveals that most violations of single-peakedness occur at the second rank while least number of violations occur at the fourth rank in treatments DB, TB, and TE.[18] This pattern helps in the following discussion of reasons for non-single-peakedness.[19]

6. Discussion and conclusion

Most studies in the experimental literature on other-regarding preferences test a particular theory or compare competing theories. In a pioneering paper, Andreoni and Miller (2002) asked a far more general question: "Can subjects' concerns for altruism or fairness be expressed in the economists' language of a well-behaved preference ordering?" They elicited choices of subjects in a series of dictator games with varying pie sizes and prices of giving to another subject and tested whether choices are consistent with the Generalized Axiom of Revealed Preference (Afriat, 1967; Varian, 1982). Although choices made by subjects

[18]These results also hold in treatments DM and TM and are described in the Supplement.

[19]It also helps in allaying a potentially important concern with the choice of probabilities in our experiment. The probability of the first, second, third, fourth, or fifth ranked alternative being used to determine the payoffs was 0.50, 0.30, 0.15, 0.05, and 0.00, respectively. It could be argued that a probability of 0.05 is 'too close' to a probability of 0.00 and this could contribute heavily to non-single-peakedness. To elaborate the concern, consider a subject in treatment DB. Suppose his true ranking – in terms of the amounts given to other – is the single-peaked ranking $9 \succ 7 \succ 4 \succ 2 \succ 0$. It is possible that the subject instead reports the non-single-peaked ranking $9 \succ 7 \succ 4 \succ 0 \succ 2$ because he treats the probability of 0.05 to be effectively 0.00. Such a concern does not seem to materialize in the data.

were quite heterogenous, almost all subjects behaved *as if* they were maximizing a utility function representing a well-behaved preference ordering.

We draw upon the work of Cox, Friedman, and Sadiraj (2008) to extend the analysis of the question raised by Andreoni and Miller. As understanding the structure of the underlying preference ordering is the main motivation, we devise a mechanism to directly elicit preference orderings. Well-behaved orderings over sure alternatives specifying sure payoffs to *self* and *other* are single-peaked.

Among subjects labeled as neoclassical about 16% exhibit non-single-peaked rankings in the pooled data. The corresponding fraction is nearly 32% among subjects labeled non-neoclassical. The fraction of subjects that the existing analytical framework can not accommodate rises from 21% to 45% with a seemingly minor change to the set of alternatives when $(9, 11)$ in treatment TB is replaced with $(10, 10)$ in TE. As our treatments are, arguably, based on the simplest possible laboratory games we believe our results provide a conservative estimate of non-single-peakedness. Nonetheless, we want to primarily focus on the qualitative patterns in the data that are summarized below.

[**F1**] Single-peakedness is significantly lower is treatment TE which contains alternative $(10, 10)$ relative to treatment TE which instead contains $(9, 11)$.

[**F2**] Positively and/or negatively other-regarding subjects are less likely to reveal single-peaked preferences in all treatments. But, the significant difference in single-peakedness across TE and TB cannot be attributed to an increase in positively-other regarding behavior in TE.

[**F3**] Non single-peakedness arises for the first time relatively more towards the top rather than the bottom of subjects rankings irrespective of the first-ranked alternative.[20]

[**F4**] There is no difference across men and women in terms of the likelihood of single-peakedness in any treatment.

[**F5**] There is no difference in the single-peakedness of dictators and comparable trustees (in treatments DB vs. TB and DM vs. TM).

In the following, we delineate the potential reasons for violations of single-peakedness, discuss the challenges involved in systematically exploring them in future studies, and highlight some potential directions for theoretical modeling of other-regarding behavior.

[20]Non-single-peakedness can arise for the first time at the second to $(n-1)^{th}$ rank for $n \geq 4$.

6.1. Reasons for violations of single-peakedness

Outcome-based other-regarding preference theories model the preferences of an individual over sure alternatives, where each sure alternative specifies sure material payoffs for *self* and *other*. Preferences over sure alternatives are modeled with the help of *one* binary relation. The binary relation is assumed to satisfy a set of core axioms. From a *theoretical* perspective, three potential sources of non-single-peakedness are as follows.

[S1] Preferences over sure alternatives cannot be modeled with only one binary relation (perhaps, because there is a mismatch between the way we formally define the set of alternatives and the way subjects mentally represent it).

[S2] Even if preferences over sure alternatives can be modeled with only one binary relation, one or more of the core preference axioms are violated.

[S3] Preferences over sure alternatives can be modeled with only one binary relation satisfying the core axioms, but subjects commit random errors.

Violations of single-peakedness may only be partly attributed to [S3] because of the patterns regarding *who* violates single-peakedness and *where* single-peakedness is violated in subjects' rankings as summarized in findings [F2] and [F3]. It seems unsatisfactory to dismiss patterns of behavior that a particular framework can not accommodate as errors in the decision making process. Hence, we shall disregard [S3] as a significant source of non-single-peakedness in the remainder of our discussion.

6.1.1 [S1] *versus* [S2]

Our experiment does not allow us to pin-point [S1] or [S2] as the unambiguous source of any particular violation of single-peakedness. However, we shall argue that the observed patterns in violations of single-peakedness suggest that all violations of single-peakedness can potentially be explained by appealing to [S1]. In contrast, it seems difficult to do so by appealing to [S2].

To see the *behavioral* role of [S1], suppose a subject in the DB treatment reports the non-single-peaked ranking $9 \succ 0 \succ 2 \succ 4 \succ 7$ in terms of the amounts given to *other*. It is possible that the subject uses two 'rationales' (in a sequential manner) to come up with this ranking (Tadenuma, 1998; Kalai, Rubinstein, and Spiegler, 2002; Manzini and Mariotti, 2007). The first rationale focusses on identifying the most 'generous' alternative and ranking it first while the second rationale involves focussing on one's own payoff thereafter.

We find it difficult to seriously question any of the core preference axioms if we assume that preferences over sure alternatives can indeed be modeled with only one binary relation. Hence, we do not regard [S2] as the fundamental source of non-single-peakedness. Consider finding [F1] which summarizes that single-peakedness drops significantly in treatment TE relative to treatment TB. Suppose, one (or more) of the core axioms gets violated significantly more often in TE which leads to the significant reduction in single-peakedness. Since the only difference between TB and TE is that alternative $(9, 11)$ has been replaced with alternative $(10, 10)$ the reason must be linked to this change.

We could not come up with a compelling argument as to why the core preference axiom(s) would fail significantly more often in TE *without* appealing to [S1] (implicitly or explicitly). Explaining the patterns regarding who violates single-peakedness (finding [F2]) and where it is violated (finding [F3]) by appealing to increased violations of one or more preference axioms without appealing to [S1] seems even more challenging.

In contrast, [S1] can, in principle, be used to provide an explanation for findings [F1], [F2], and [F3]. We first explain [F1] – the significant difference in the extent of single-peakedness across TB and TE. We do so by adapting the arguments put forth by Sen (1993), McFadden (1999), and Kalai, Rubinstein, and Spiegler (2002) in a related context.

Consider a trustee in treatment TB faced with the task of ranking the alternatives which involve giving 0, 2, 4, 7, and 9 to *other*. Most subjects might interpret these alternatives 'uni-dimensionally' by noting that they involve giving different amounts to *other*. The presence of the equal-split alternative in treatment TE, perhaps, disturbs this uni-dimensional interpretation of the alternatives. A significant fraction of subjects may think that giving 10 not only amounts to giving the most, but it is the 'fair' thing to do as well. This informal description can be formally captured by allowing a higher-dimensional description of alternatives or by modeling preferences with the help of multiple binary relations.[21]

[S1] is fully consistent with patterns regarding who violates single-peakedness (finding [F2]) as discussed in the second paragraph of this sub-section. Finally, [S1] is also consistent with patterns regarding where single-peakedness is violated (finding [F3]). If a subject uses two rationales with the first one focussed on generosity and the second one focussed on payoffs to self, then what we refer to non-single-peakedness because *generosity runs out*, can be explained. Similarly, non-single-peakedness when *selfishness runs out* could be due to the possibility that a subject uses the same two rationales but in the reverse order.[22]

[21]We have specified the set of alternatives to be a finite subset of $\mathbb{R}^2_+$. The language in which a subject describes the set of alternatives to himself may be such that the set of alternatives ceases to be a subset of $\mathbb{R}^2_+$. The interested reader may refer to Kalai, Rubinstein, and Spiegler (2002) for a discussion of why relying on multiple binary relations is an appropriate way of modeling such situations.

[22]It need not be the case that the first rationale is utilized by the subject to determine only the first-ranked alternative.

6.2. Methodological issues

Given that we use a probabilistic mechanism to ultimately infer the preferences over sure alternatives, it is necessary to discuss the potential concerns introduced by the use of such a mechanism. Preferences over the sure alternatives are extended to preferences over lotteries (whose potential outcomes are the sure alternatives) by using some theory of decision making under risk (which almost always happens to be expected utility theory in applications). Thus, from a theoretical perspective, violations of single-peakedness in our experiment can also be attributed to the following source.

[S4] Even if preferences over sure alternatives can be modeled with one binary relation that satisfies the core axioms, preferences over lotteries are such that the reported ranking can not be unambiguously interpreted as the preference ranking over the sure alternatives. (see [Q1] in Section 3.2.2.).

Suppose the true preference ranking of a subject over the sure alternatives in treatment DB is $9 \succ 7 \succ 4 \succ 2 \succ 0$. Further suppose that his preferences over lotteries can not be modeled using any theory of decision making under risk which leads to an affirmative answer to [Q1] in Section 3.2.1. In such a case, the subject could report the non-single-peaked ranking $9 \succ 0 \succ 2 \succ 4 \succ 7$ instead of his true ranking but it would be inaccurate to conclude that the subject's ranking over the sure alternatives is not in line with the existing theories of other-regarding preferences.

Yet, we believe that the choice of a probabilistic mechanism, at best, only partly explains non-single-peakedness. Recall the comparisons between treatment TE which contains the equal-split and treatment TB which does not. We find (i) differences in the distribution of first-ranked alternatives and (ii) in the extent of single-peakedness across these two treatments. Gueth et al. (2001) report results from mini-ultimatum games with and without the equal-split. The proposer could propose $(10, 10)$ or $(3, 17)$ in their 'equal-split' treatment (with the first entry referring to the payoff for the recipient). In the 'no-equal-split' treatment, the proposer could propose $(9, 11)$ or $(3, 17)$. Alternative $(3, 17)$ is proposed significantly more often in the 'no-equal-split' treatment. Similar results are obtained by Falk, Fehr, and Fischbacher (2003).[23] It is crucial to note that these findings are similar to (i) and the combination of (i) and (ii) can not be explained by appealing to [S4] because Gueth et al. (2001) and Falk et al. (2003) elicit choices of proposers using a

[23] The experimental literature on bargaining contains several studies that highlight the role of the availability of equal-split. In a recent study Anbarci and Feltovich (2011) find that responsiveness to changes in the disagreement outcome is relatively higher when one agent has a disagreement payoff above half the pie size (so that 50-50 splits are not individually rational).

non-probabilistic mechanism. In contrast, a reasoning along the lines of [S1] can explain these findings as the presence of the equal-split alternative can induce a significant fraction of subjects to cognitively represent the alternatives in a multi-dimensional way.

Unfortunately, no currently known mechanism seems capable of unambiguously distinguishing between [S1] and [S4] as the source a particular violation of single-peakedness. To see why, consider an alternative way of implementing the treatment DB. The dictator can be asked to make a choice from all the subsets of the set of alternatives that contain at least two alternatives. One of the choice problems can be selected randomly for payment with all problems being equally likely to be selected. The choices made by the dictator can be used to check whether or not the inferred ranking over the alternatives is single-peaked. One would not be able to pin-point the precise reason for violations of single-peakedness even with this design. [S1] clearly qualifies as a potential reason. [S4] can also be a potential reason because such a design will provide subjects the incentive to truthfully report their preferred choice across all choices problems only if they behave in line with expected utility theory (Cox, 2010).[24] Devising a way to distinguish between potential sources of non-single-peakedness remains a non-trivial open problem.

6.3. Concluding remarks

The question raised by Andreoni and Miller (2002) exemplifies that the literature on other-regarding preferences aims to go beyond *as if* explanations for observed choices (Fehr and Falk, 2002; Blanco, Engelmann, and Norman, 2011) and gain a better understanding of the underlying preferences. The goal of this literature makes it pertinent to devise ways to chip away at the *as if* approach and directly investigate the preference structure.[25] Given the enormity of this task, eliciting rankings is a small step in this direction. Nonetheless, it provides significantly more information about the structure of individual preferences at no additional monetary cost to the experimenter since the first-ranked alternatives are quite likely to serve as a very good proxy for the choice subjects would have made if they were asked to choose an alternative. The idea of single-peakedness and the MDP mechanism can potentially be utilized to explore issues related to preference structures beyond the domain of other-regarding preferences.

[24]The interested reader may refer to Cox, Sadiraj, and Schmidt (2011) for a discussion of several payment mechanisms for experiments where subjects make multiple-decisions. Subjects have to choose between lotteries in their study. However, concerns regarding wealth effects and whether subjects treat each decision task as a separate one are shared by experiments where subjects make multiple decisions but each decision involves choosing one sure alternative from a set of sure alternatives. In theory, none of the payment mechanisms solves the problem to pin-pointing the exact source of violations of single-peakedness if subjects are asked to make multiple decisions.

[25]It is often the case that choices made by subjects are used to estimate the parameters of a flexible utility function (Andreoni and Miller, 2002; Fisman, Kariv, and Markovits, 2007). Such an exercise would become significantly more meaningful if the underlying preferences are indeed single-peaked.

Fudenberg and Levine (2011) prove that defining other-regarding preferences only over sure alternatives and extending them to lotteries in line with expected utility theory involves a fundamental conflict between the IIA axiom and behaviorally appealing notions of ex-ante fairness.[26] This suggests interpreting theoretical predictions of strategic interactions between agents having other-regarding preferences with caution. because analyzing strategic interactions between agents with other-regarding preferences necessarily involves extending preferences over sure alternatives to preferences over lotteries. As argued by Fudenberg and Levine (2011), a fruitful direction for future research would be to devise other-regarding preference structures that are behaviorally sound and can be extended from sure alternatives to lotteries in empirically meaningful ways.

We conclude by noting that recent advances in individual choice theory suggest that a framework which models preferences using multiple binary relations may provide a better understanding of individual behavior (Manzini and Mariotti, 2007). Such an approach might be an interesting avenue for future theoretical and experimental research on other-regarding preferences as it is a natural extension of the standard framework (Tadenuma, 1998; Kalai, Rubinstein, and Spiegler, 2002; Manzini and Mariotti, 2007). In addition, it could provide a route to formalize the suggestion that individuals might be using a family of rules of thumb or heuristics (Cox, 2004; Wilson, 2010).

References

[] Afriat, S. (1967). "The Construction of a Utility Function From Expenditure Data." *International Economic Review*, 8, 67-77.

[] Anbarci, N., and N. Feltovich. "How Responsive are People to Changes in their Bargaining Position? Earned Bargaining Power and the 50–50 Norm." *Working Paper.*

[] Andreoni, J., and J. H. Miller. (2000). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 70, 737-753.

[] Andreoni, J., and L. Vesterlund. (2001). "Which is the Fair Sex? Gender Differences in Altruism." *Quarterly Journal of Economics*, 116, 293-312.

[] Andreoni, J., M. Castillo, and R. Petrie. (2003). "What do Bargainers' Preferences Look Like? Exploring a Convex Ultimatum Game." *American Economic Review*, 93, 672-685.

[] Bardsley, N. (2008). "Dictator Game Giving: Altruism or Artefact?" *Experimental Economics*, 11, 122-133.

[] Berg, J., J. Dickhaut, and K. McCabe. (1995). "Trust, Reciprocity and Social History."

[26] The interested reader may refer to Fudenberg and Levine (2011) for details. There is a close parallel between their axioms of 'ex-ante fairness for me' and 'ex-ante fairness for you' and what we refer to as 'generosity runs out' and 'selfishness runs out,' respectively. An attempt to directly test the axioms proposed by Fudenberg and Levine should be undertaken with the caution since it might not be easy to pin-point whether (and when) [S1] or [S4] explains the results.

Games and Economic Behavior, 10, 122-142.

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

[] Black, D. (1958). *Theory of Committees and Elections.* Cambridge University Press, Cambridge.

[] Blanco, M., D. Engelmann, and H. T. Normann. (2011). "A Within-Subject Analysis of Other-Regarding Preferences." *Games and Economic Behavior*, 72, 321-338.

[] Camerer, C. F., and K. Weigelt. (1988). "Experimental Tests of a Sequential Equilibrium Reputation Model." *Econometrica*, 56, 1-36.

[] Casari, M. and T. N. Cason. (2009). "The Strategy Method Lowers Measured Trustworthy Behavior." *Economics Letters*, 103, 157-159.

[] Chakravarty, S., Y. Ma, and S. Maximiano. (2011). "Lying and Friendship." *Working Paper.*

[] Charness, G., and M. Rabin. (2002). "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117, 817-869.

[] Cox, J. C. (2004). "How to Identify Trust and Reciprocity." *Games and Economic Behavior*, 46, 260-281.

[] Cox, J. C., D. Friedman, and V. Sadiraj. (2008). "Revealed Altruism." *Econometrica*, 76, 31-69.

[] Cox, J. C. (2010). "Some Issues of Methods, Theories, and Experimental Designs." *Journal of Economic Behavior and Organization*, 73, 24-28.

[] Cox, J. C., V. Sadiraj, and U. Schmidt. (2011). "Paradoxes and Mechanisms for Choice under Risk." *Working Paper.*

[] Craven, J. (1992). *Social Choice.* Cambridge Univ. Press, Cambridge, 1992.

[] Croson, R., and U. Gneezy. (2009). "Gender Differences in Preferences." *Journal of Economic Literature*, 47, 1-27.

[] Dufwenberg, M., and G. Kirchsteiger. (2004). "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47, 268-298.

[] Falk, A. E. Fehr, and U. Fischbacher. (2003). "On the Nature of Fair Behavior." *Economic Inquiry*, 41, 20-26.

[] Falk, A., and U. Fischbacher. (2006). "A Theory of Reciprocity." *Games and Economic Behavior*, 54, 293-315.

[] Fehr, E., and A. Falk. (2002). "Psychological Foundations of Incentives." *European Economic Review*, 46, 687-724.

[] Fehr, E. and K. M. Schmidt. (1999). "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 817-868.

[] Fischbacher, U. (2007). "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics*, 10, 171-178.

[] Fisman, R., S. Kariv, and D. Markovits. (2007). "Individual Preferences for Giving." *American Economic Review*, 97, 1858-1876.

[] Forsythe, R., J. Horowitz, N. Savin, and M. Sefton. (1994). "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, 6, 347-369.

[] Friedman, M. (1953). *Essays in Positive Economics*, University of Chicago Press, Chicago.

[] Fudenberg, D., and D. K. Levine. (2011). "Fairness, Risk Preferences, and Independence: Impossibility Theorems." *Working Paper*. [] Güth, W., R. Schmittberger, and B. Schwarze. (1982). "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, 3, 367-388.

[] Güth, W., S. Huck, and W. Müller. (2001). "The Relevance of Equal Splits in Ultimatum Games." *Games and Economic Behavior*, 37, 161-169.

[] Kahneman, D., and A. Tversky. (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47, 263-291.

[] Kalai, G., A. Rubinstein, and R. Spiegler. (2002). "Rationalizing Choice Functions by Multiple Rationales." *Econometrica*, 70, 2481-2488.

[] List, J. A. (2007). "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy*, 115, 482-493.

[] Manzini, P., and M. Mariotti. (2007). "Sequentially rationalizable Choice." *American Economic Review*, 97, 1824-1838.

[] McFadden, D. L. (1999). "Rationality for Economists." *Journal of Risk and Uncertainty*, 19, 73-105.

[] Morgenstern, O., and J. V. Neumann. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.

[] Quiggin, J. (1982). "A Theory of Anticipated Utility." *Journal of Economic Behavior and Organization*, 3, 323-343.

[] Sen, A. (1993). "Internal Consistency of Choice." *Econometrica*, 61, 495-521.

[] Tadenuma, K. (1998). 'Efficiency First or Equity First? Two Principles and Rationality of Social Choice." *Journal of Economic Theory*, 104, 462-472.

[] Varian, H. R. (1982). "The Nonparametric Approach to Demand Analysis." *Econometrica*, 50, 945-972.

[] Wilson, B. J. (2010). "Social Preferences aren't Preferences." *Journal of Economic Behavior and Organization*, 73, 77-82.

[] Yaari, M. E. (1987). "The Dual Theory of Choice Under Risk." *Econometrica*, 55, 95-115.