Kerschbamer, Rudolf

Working Paper

The geometry of distributional preferences and a non-parametric identification approach

Working Papers in Economics and Statistics, No. 2013-25

Provided in Cooperation with:
Institute of Public Finance, University of Innsbruck

Suggested Citation: Kerschbamer, Rudolf (2013) : The geometry of distributional preferences and a non-parametric identification approach, Working Papers in Economics and Statistics, No. 2013-25, University of Innsbruck, Research Platform Empirical and Experimental Economics (eeecon), Innsbruck

This Version is available at:
https://hdl.handle.net/10419/101108

University of Innsbruck

The geometry of distributional preferences and a non-parametric identification approach

Rudolf Kerschbamer

Working Papers in Economics and Statistics

2013-25

Research Platform
eeecon
Empirical and Experimental Economics
University of Innsbruck
http://eeecon.uibk.ac.at/

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact Address:
University of Innsbruck
Department of Public Finance
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

The Geometry of Distributional Preferences and a Non-Parametric Identification Approach[*]

Rudolf Kerschbamer

Department of Economics, University of Innsbruck[#]

This Version: October 2013

Abstract: This paper proposes a geometric delineation of distributional preference types and a non-parametric approach for their identification in a two-person context. It starts with a small set of assumptions on preferences and shows that this set (i) naturally results in a taxonomy of distributional archetypes that nests all empirically relevant types considered in previous work in economics and social psychology; and (ii) gives rise to a clean experimental test design that discriminates between archetypes according to core features of preferences rather than properties of specific modelling variants. As a by-product the test yields a two-dimensional index of preference intensity.

JEL Classifications: C90, D63, D64, C81, B41

Keywords: Distributional preferences, social preferences, other-regarding preferences, social value orientations, behavioral economics, experimental economics

[*] A previous version of this paper circulated under the title "A Parsimonious Experimental Test to Identify Type and Intensity of Distributional Preferences at the Individual Level".

[#] Universitätsstrasse 15, A-6020 Innsbruck, Austria; e-mail: Rudolf.Kerschbamer@uibk.ac.at; phone: ++43 512 507 7400

"Everything should be made as simple as possible, but not one bit simpler."

attributed to Albert Einstein (1879-1955**)**

1 Introduction

Many economists' default assumption is that all agents are exclusively motivated by their own material self-interest. This assumption is in sharp contrast to both day-to-day experience and empirical evidence gathered by psychologists and experimental economists in the last decades. This has led researchers to propose theories of other-regarding preferences, where arguments beyond material self-interest enter the decision maker's utility function. Typical examples of such arguments are other people's (material) well-being (as in distributional preferences models),[1] others' opportunities and expected or observed behavior (as in reciprocity models),[2] others' payoff expectations (as in guilt aversion models),[3] or others' other-regarding concerns (as in type based models).[4]

The present paper focuses on the first of the above mentioned subclasses, i.e. on distributional preferences, where besides one's own material payoff the (material) well-being of others enters an agent's utility function. Distributional preferences have been shown to be behaviorally relevant in important market and non-market environments – see Sobel (2005) and Fehr and Schmidt (2006) for excellent surveys. The current paper adds to this literature by proposing a geometric delineation of distributional preference types and a *simple* non-parametric approach for their identification in a two-person context.

[1]The major distributional types discussed in the literature are altruism (Becker 1974, Andreoni and Miller 2002) and surplus maximization (Engelmann and Strobel 2004); inequality aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000), difference aversion (Charness and Rabin 2002) and egalitarian motives (Dawes et al. 2007, Fehr et al. 2008); maximin (Engelmann and Strobel 2004), Rawlsian (Charness and Rabin 2002) and Leontief preferences (Andreoni and Miller 2002, Fisman et al. 2007); spiteful preferences (Levine 1998) and concerns for relative income (Duesenberry 1949); envy (Bolton 1991, Kirchsteiger 1994, Mui 1995); and equity aversion (Charness and Rabin 2002, Fershtman et al. 2012).

[2] Positive (negative) reciprocity is the propensity to repay observed or expected generous (mean) actions of others by choosing actions that are generous (mean) to others. See Rabin (1993), Fehr et al. (1998), Charness and Rabin (2002), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Bolle and Kritikos (2006), Cox, Friedman, and Gjerstad (2007) and Cox, Friedman, and Sandiraj (2008) for theoretical models, and Fehr and Gächter (2000a) for an overview of experimental evidence and of implications of reciprocity.

[3] In guilt aversion models people have a disposition to feel guilty when letting down others' payoff expectations implying that an agent's benevolence towards another person depends on what he believes that the other expects to receive. See Charness and Dufwenberg (2006) and Battigalli and Dufwenberg (2007).

[4] In type-based models the weight placed on the material payoff of another player depends on one's own (distributional) basic attitude as well as the perceived basic attitude of the other player. For instance, in Levine (1998) people are altruistic (or spiteful) to various degrees and –given their own basic altruism– they place more positive (negative) weight on the money received by an opponent who is believed to be more altruistic (spiteful). Andreoni and Bernheim (2009) have a similar model in which others' other-regarding concerns are evaluated in terms of inequality aversion (and not in terms of altruism).

Identification of distributional preferences has been the topic of numerous papers – see Section 4 for a review of the literature. These pioneering studies – which have greatly advanced our understanding of non-selfish behavior – suffer from at least one of two methodological shortcomings. First, the tests employed typically discriminate between the members of a somewhat arbitrary list of distributional types; and secondly, the identification procedures typically rely on strong structural assumptions.[5]

Regarding the former dimension – the *set of distributional types tested for* – previous studies either start with a given list of types, or they employ a test design that allows discriminating only between the members of a limited set of types.[6] For instance, the path-breaking dictator-game study by Andreoni and Miller (2002) distinguishes between *selfish, Leontief* and *perfect substitutes* preferences, plus weak incarnations of those types; the follow up study by Fisman et al. (2007) employs a richer design and discriminates between *self-interested, lexself* (lexicographic for self over other), *social welfare* and *competitive* types plus some mixes thereof; the pioneering discrete choice study by Engelmann and Strobel (2004) tries to disentangle *efficiency concerns* (defined as surplus maximization), *maximin* preferences and (two modelling variants of) *inequality aversion*; Blanco et al. (2011) discriminate between *selfish* and various intensities of piecewise linear *inequality aversion*; Charness and Rabin (2002), Cabrales et al. (2010) and Iriberri and Rey-Biel (2013) allow for *self-interested, social welfare, difference-averse* and *competitive* preferences; and the ring-test – originally developed by social psychologists to assess "social value orientations"[7] and recently used by economists to identify type and intensity of distributional concerns[8] – discriminates between *altruists, cooperators, individualists, competititors, aggressors, martyrs, masochists and sadomasochists.*

Turning to the second dimension – the *structural assumptions imposed* – the identification procedures employed in previous studies typically rely on strong assumptions regarding the form of the utility or motivational function meant to represent preferences. For instance, the ring-test is based on the assumption of *linear preferences*; the studies by Cabrales

[5] One might argue that the list of distributional types tested for in the current paper is to some degree arbitrary, too. This is correct, of course. As will become clear below, the main innovation of the present paper in this regard is to derive the number and core properties of types from a small set of primitive assumption on preferences. This is in contrast to previous studies which either start with a given list of types or a specific model of preferences.

[6] From the papers mentioned in the main text Charness and Rabin (2002), Engelmann and Strobel (2004), Cabrales et al. (2010), Blanco et al. (2011) and Iriberri and Rey-Biel (2013) are examples of the former track (starting with a given set of types and designing tests to discriminate between the members of the set; or starting with a functional form and estimating the free parameters), while Andreoni and Miller (2002) and Fisman et al. (2007) are examples of the latter track (starting with a test without specifying a priori which types are tested for).

[7] Standard references are Griesinger and Livingston (1973) and Liebrand (1984).

[8] See, for instance, Offerman et al. (1996), Sonnemans et al. (1998), van Dijk et al. (2002), Brosig (2002), Brandts et al. (2009), or Sutter et al. (2010).

et al. (2010), Blanco et al. (2011) and Iriberri and Rey-Biel (2013) employ identification procedures based on the piecewise linear model originally introduced by Fehr and Schmidt (1999) as a description of self-centered inequality aversion and later extended by Charness and Rabin (2002) to allow for other forms of distributional concerns and thereby assume *piecewise linearity*; and Andreoni and Miller (2002), Fisman et al. (2007) and Cox and Sadiraj (2012) check consistency with – and estimate parameters of – standard or modified constant elasticity of substitution (*CES*) *utility functions*.

Summing up the above discussion we conclude (i) that there is neither an agreement in the literature on what the relevant set of distributional basic motivations – defined as the manner in which people care about the (material) well-being of others – is, nor on how to delimitate distributional types; and (ii) that existing studies employ identification procedures that rely on strong structural assumptions as, for instance, linearity, piecewise linearity or standard or modified CES forms. By using a systematic approach based on a small set of primitive assumptions on preferences, the present paper offers an improvement in both dimensions. It shows (i) that this set of assumptions naturally results in a well delineated, mutually exclusive and comprehensive distinction between nine archetypes of distributional concerns; and (ii) that this set gives rise to a simple non-parametric experimental test that discriminates between the archetypes according to core features of preferences rather than properties of specific modeling variants or functional forms. As a byproduct the test yields a two-dimensional index of preference intensity.

While the primary purpose of this paper is methodological, the experimental results obtained in an implementation of the test also produce some substantive insights. For instance, the result that – consistent with the theoretically appealing assumption that distributional preferences are convex – about 95% of the subjects reveal (weakly) more benevolent (less malevolent) preferences in the domain of advantageous than in the domain of disadvantageous inequality. A second interesting detail is that beyond selfish subjects, the empirically most frequent distributional archetypes are those who exhibit (at least weakly) positive attitudes towards others in both domains (i.e., altruism and maximin), while archetypes that imply a negative attitude in at least one of the domains are empirically by far less important (the behavior of less than a forth of the subjects is consistent with any form of inequality aversion, for instance, and the choices of less than 7% of the subject population are consistent with spite).[9]

The rest of the paper is organized as follows: Section 2 starts by introducing the assumptions on which the analysis is based and argues that those assumptions are fulfilled by

[9] The finding that there are very few malevolent people in variants of the dictator game is in line with previous studies – see Charness and Rabin (2002) or Engelmann and Strobel (2004), for instance.

all major modeling variants of distributional preferences discussed in the economic literature. It continues by highlighting the core features of different distributional preference types discussed in the literature and shows that delimiting types according to the proposed assumptions naturally leads to a distinction between nine mutually exclusive and comprehensive archetypes of distributional concerns. Section 3 introduces the identification procedure (the "*XY*-test"). It starts with the intuition behind the proposed identification approach (Subsection 3.1) and then presents the basic version of the test (Subsection 3.2). Furthermore, a two-dimensional index for identifying the archetype and characterizing the intensity of distributional concerns – the (*x, y*)-score – is introduced, and a graphical representation of the type-intensity distribution is proposed (in Subsection 3.3). The section ends with a discussion of several extensions and modifications of the test (in Subsection 3.4). Section 4 highlights the main differences between the current approach and other tests proposed in the literature. Section 5 relates the (*x, y*)-score to other measures of type and intensity of distributional concerns and discusses the pros and cons of replacing it by a cardinal metric (like the willingness to pay for changes in the income of the other). Section 6 illustrates the working of the identification procedure by reporting experimental results generated with the symmetric basic version of the test, and Section 7 concludes. Implementation issues for the case where the test is used as a tool in experimental economics (to address research questions in which distributional preferences play an important role, to control for subject pool effects, or to help to interpret data from other unrelated experiments) are discussed in Appendix A, while Appendix B contains a more detailed discussion of the ring-test, probably the closest competitor to the current approach when intended as an add-on to other experiments. Appendix C contains the instructions of the experiment reported in Section 6.

2 Assumptions and Core Features of Different Types of Distributional Preferences

2.1 The Three Basic Assumptions

Let $a = (m, o)$ denote an income allocation that gives material payoff m (for "my") to the decision maker (DM or "agent") and material payoff o (for “other”) to the other person. Then we require:

Assumption 1 (ordering and continuity): Admissible preference relations on income allocations (m, o) can be represented by a continuous utility or motivational function $u(m, o)$ that assigns a real-valued index to all $(m, o) \in R^2$.

In terms of axioms on preferences Assumption 1 requires *ordering* (completeness and transitivity) *and continuity*. While ordering is important for the arguments below (as it is for

substantial parts of economic theory), continuity is not.[10] The second assumption requires that – holding the material payoff of the other person constant – the DM's utility is strictly increasing in her own material payoff:

Assumption 2 (strict *m*-monotonicity): Holding *o* constant, the agent's well-being is strictly increasing in *m*. That is, $\partial u/\partial m > 0$ for all $(m, o) \in R^2$.

Strict *m*-monotonicity is quite a natural assumption. It is violated, for instance, if the DM is willing to burn her own monetary payoff because she feels bad whenever she has (much) more than the other person. Such behavior is essentially never observed in experiments. The third assumption requires that the DM's general attitude towards the other person (i.e., whether she is benevolent, neutral, or malevolent to the other) depends only on whether the other person has more or less monetary payoff than the DM herself:

Assumption 3 (piecewise *o*-monotonicity): The general attitude of the agent toward the other person (i.e., whether an increase in *o* has a positive, a zero, or a negative impact on the agent's well-being) depends only on whether the agent is ahead ($m > o$) or behind ($m < o$). That is, $\text{sign}(\partial u/\partial o)$ depends (only) on $\text{sign}(m - o)$ for all $(m, o) \in R^2$.

Piecewise *o*-monotonicity is both permissive and restrictive, depending on the perspective. It is permissive because it allows for all major variants of distributional preferences that have been discussed in the economic literature – see the discussion at the end of this and in the next subsection. Piecewise *o*-monotonicity is also restrictive because it implies (i) that preferences only depend on outcomes, not on the way they are achieved (this is the defining feature of distributional preferences); and (ii) that the reference point for the evaluation of allocations (if one is used) is an equal-material-payoffs allocation.

Ad (i) The implication that preferences only depend on outcomes is likely to be violated in many important applications. For instance, in strategic interactions (where the other person has an opportunity to move and thereby a possibility to influence the payoff of the DM) *beliefs about intentions behind observed or expected action choices* of the other person potentially play a role (see the literature on reciprocity and related concepts cited in Footnote 2). Also, in some games *beliefs about the payoff expectations of the other person* seem to influence behaviour (see the literature on guilt aversion and related concepts cited in Footnote 3). Furthermore, in a richer environment, where agents have more information on each other,

[10] Continuity simplifies the presentation of the assumptions and the description of the core features of different archetypes of distributional preferences but is neither needed nor used in any other part of the paper. Continuity is not needed because the identification procedure proposed here uses information on the shape of revealed upper and lower contour sets to derive bounds on indifference sets. So, from a theoretical point of view, the procedure works even if indifference sets are singletons (as is the case for "lexself" preferences discussed by Fisman et al. 2007, for instance).

beliefs about the other-regarding concerns of the other person may play a role (as in the literature on type-based models cited in Footnote 4). Finally, *features of the situation* (such as context, entitlements, properties of the outcome generating process, etc.) *or the DM* (such as a code of conduct, or a preference for honesty) might shape behaviour. Knowing that all those factors might be behaviourally relevant in a richer environment it seems important that distributional preferences are identified in a non-strategic setting and a neutral frame to avoid confounds. This is not to say that distributional preferences are unimportant in richer environments, of course, but rather that they cannot be unambiguously identified there.

Ad (ii): Some distributional archetypes discussed in real life and in the literature (most importantly, inequality aversion and egalitarian motives; maximin, Rawlsian and Leontief preferences; and envy) are inevitably defined in terms of a "reference location" (an interval, a point, or whatsoever), where the DM's general attitude towards the other changes sign. In theory, this reference location can be anything, of course, and it can differ among individuals. In existing models of reference-dependent distributional concerns, the reference location is a point, and the point is the egalitarian one for all individuals (see, for instance, Bolton 1991, Mui 1995, Fehr and Schmidt 1999, Bolton and Ockenfels 2000, or Charness and Rabin 2002). While Assumption 3 is more agnostic than existing models of reference-dependent distributional concerns, it is still restrictive.[11] For instance, there might exist individuals who consider it fair to get 20% more than others but unfair to get 30% more. Assumption 3 does not allow for this. While it would be feasible, in principle, to generalize Assumption 3 (and the test relying on it) so as to allow for heterogeneous reference points, this would seriously impair simplicity and transparency: Ultimately the aim of the paper is to propose a classification of subjects in distributional preference types that is helpful in organizing experimental data. For that purpose we need some kind of clustering and not a different distributional type for each single individual. Stated differently, as *any* model the approach proposed here is by design an abstraction of realty, and hence is deliberately constructed so as to *not* explain some behavior, in return for parsimony.

While parsimony calls for a unique reference point, it does not suggest equality as the reference point. Equality is suggested (a) by normative considerations and (b) by empirical evidence.

Regarding the *normative basis* Konow (1993, p.1194) argues in his review of theories of justice that "[t]he most primitive, and probably oldest, notion of justice associates equity

[11] What Assumption 3 essentially requires is that the egalitarian outcome is somehow focal among those subjects who change their general attitude (i.e., whether they are benevolent, neutral or malevolent) towards others at some point. It does not require the attitude to change, though. In other words, while existing models of reference-dependent distributional concerns assume that "*something special happens at equality*", Assumption 3 "only" requires that "*if something special happens with preferences then it happens around equality*".

with equality." And indeed, two of the most prominent theories of distributional justice – Egalitarism and Rawlsianism– have explicitly been formulated with reference to equality. Rawls (1971), for instance, claims that two equity principles would be chosen in the 'original position' (that is, 'behind a veil of ignorance'), first the equality principle asking for equal rights and opportunities, and second the difference principle demanding that goods are distributed equally unless an unequal distribution is to the advantage of the least favoured. This is nothing else than a maximin rule for the distribution of resources that has equality as a benchmark. One might be inclined to ask what the Rawlsian original position has to do with the distributional concerns of subjects in the lab, or more general, with the fairness perceptions of actual people. Binmore (2005, p. 18) answers this question as follows: "All fairness norms in actual use share the deep structure of Rawls' original position. This deep structure is biologically determined, and hence universal in the human species [....]." And on another occasion (on p. 14) he argues: "Why do we care about fairness? ... [W]e care because fairness is evolution's solution to the equilibrium selection problem for our ancestral game of life. What evidence is there for this conjecture? All the societies studied by anthropologists that survived into modern times with a pure hunter-gathering economy had similar social contracts with a similar deep structure. [...] They tolerate no bosses, and they share on a very egalitarian basis." The statement that equality is evolution's solution to the equilibrium selection problem is discussible, of course, the statement that equality is suggested by normative considerations seems safe, though.

Regarding *empirical evidence* Andreoni and Bernheim (2009, 1607f) cite several studies showing that equal sharing is common in the context of joint ventures among business firms, partnerships among professionals, share tenancy in agriculture, and bequests to children. They also provide evidence indicating that equality is a frequent outcome of negotiation and conventional arbitration in the field. In lab-experiments the assumption that the egalitarian outcome is somehow focal among subjects who change their general attitude towards others at some point seems even more natural than in the field: Subjects enter the laboratory as equals, their roles are assigned randomly and they have absolutely no information about each other. It seems therefore quite plausible that those subjects who attribute special meaning to an allocation (again, nothing in Assumption 3 requires them to do so) do this to the egalitarian one. And there is indeed considerable support for this assumption in existing experimental data. For instance, one of the stylized facts in standard dictator games is precisely that a sizeable fraction of the subject population voluntarily cedes exactly half of the pie to the recipient, and that very few subjects cede more (Camerer 1997). This result survives even in experiments where the action space is continuous and where the price for giving is quite high (see Andreoni and Miller 2002, for instance). The frequency of equal divisions is even higher in ultimatum

games, where expectations about the "reference point" of the recipient enter the picture (see Camerer 2003). While all this evidence indicates that the egalitarian outcome has something special for a substantial fraction of subjects, it does not tell us anything about the exact fraction of subjects for whom this is the case.[12] But this is exactly (one of) the question(s) the proposed test tries to address.

As is easily checked almost all (modeling) variants of distributional preferences discussed in the economics literature satisfy assumptions 1-3, notable exceptions being lexself preferences (discussed by Fisman et al. 2007) which – in a strict interpretation – violate the continuity part of Assumption 1, and maximin (or Rawlsian, or Leontief) preferences (discussed by Andreoni and Miller 2002, Charness and Rabin 2002, and Engelmann and Stobel 2004, for instance) which – in their purest form (but not in the form typically discussed in the literature) – violate strict *m*-monotonicity.[13]

2.2 Nine Archetypes of Distributional Preferences and their Core Features

This subsection highlights the core features of different distributional preference types discussed in the literature and shows that delimiting types according to the three basic assumptions introduced in the previous subsection – the *ordering* part of Assumption 1, *strict m-monotonicity* and *piecewise o-monotonicity* – naturally leads to a distinction between nine mutually exclusive and comprehensive archetypes of distributional concerns.

- First consider *selfish* or *own-money-maximizing* preferences. They can be considered as a degenerated version of distributional preferences where an agent's well-being neither increases nor decreases in the monetary payoffs of other agents. Thus, the core property of selfish preferences in a two-person context is that indifference curves in (*m, o*) space are vertical (see Table 1 for the mathematical statement and Figure 1 for the geometric representation).
- The well-being of an *altruistic* agent increases in the monetary or utility payoffs of other agents (Becker 1974, Andreoni and Miller 2002); the well-being of an *efficiency loving* or *surplus maximizing* agent (Engelmann and Strobel 2004), the well-being of an agent

[12] Here note that an egalitarian subject –according to the definition given in the next subsection– does not necessarily decide for an egalitarian allocation in a dictator game: If her preferences are smooth and satisfy strict *m*-monotonicity she will rather accept some advantageous inequality as this increases the utility derived from the own money component at a low cost in terms of the second component; similar arguments hold for other reference-dependent motives and other game forms.

[13] The ERC model by Bolton and Ockenfels (2000) permits even violations of weak *m*-monotonicity. The same is true for (models of) some "social value orientation types" (the synonym for distributional preference types used by social psychologists), most notably, "martyrdom", "masochism" and "sadomasochism" (see Appendix B for details). It is important to note, however, that even in the social psychology literature violations of *m*-monotonicity are empirically irrelevant (I know of no study finding more than 5% of subjects in the mentioned categories).

with *perfect substitutes* preferences (Andreoni and Miller 2002) and the well-being of an agent with *social welfare* preferences (Charness and Rabin 2002, Fisman et al. 2007) increases in the (weighted or unweighted) sum of payoffs. In all cases, well-being increases in *o* everywhere; thus, indifference curves in (*m, o*) space are negatively sloped everywhere (if *o* increases *m* has to decrease to hold the agent indifferent).

- An agent is *spiteful* (Levine 1998), or *competitive* (Charness and Rabin 2002), or *status seeking* or *interested in relative income* (Duesenberry 1949), if her well-being decreases in the payoffs of others everywhere; so the core property of such preferences is positively sloped indifference curves in (*m, o*) space.
- The well-being of an *envious* or *grudging* agent decreases in the payoffs of agents who have more but is unaffected by the payoffs of agents who have less (the role of envy has been emphasized by Bolton 1991 and Mui 1995, for instance); thus, the core property of envious preferences is positively sloped indifference curves in the domain of disadvantageous inequality and vertical indifference curves in the domain of advantageous inequality.[14]
- The well-being of an agent with *maximin* preferences (Engelmann and Strobel 2004), *Rawlsian* preferences (Charness and Rabin 2002), or *Leontief* preferences (Andreoni and Miller 2002, Fisman et al. 2007) increases in the lowest of all agents' payoffs. Thus, its defining feature in a two-person context is that indifference curves in (*m, o*) space are negatively sloped if inequality is advantageous and vertical otherwise.
- An agent is *inequity* or *inequality averse* (Fehr and Schmidt 1999, Bolton and Ockenfels 2000), or *difference averse* (Charness and Rabin 2002, Fisman et al. 2007), or *egalitarian* (Dawes et al. 2007, Fehr et al. 2008) if she incurs a disutility when other agents have either higher or lower payoffs (as in the model by Fehr and Schmidt 1999), or when the agent's payoff differs from the average payoff of all agents (as in Bolton and Ockenfels 2000). Consequently, the defining feature of inequality averse or egalitarian preferences in a two-person context is negatively sloped indifference curves in the domain of advantageous and positively sloped indifference curves in the domain of disadvantageous inequality.
- The opposite constellation, benevolence in the domain of disadvantageous inequality combined with malevolence in the domain of advantageous inequality, is referred to as *equality aversion* (by Hennig-Schmidt 2002, for instance), or as *equity aversion* (e.g. by Charness and Rabin 2002 and by Fershtman et al. 2012). Its defining feature in a two-

[14] Envy has also been discussed by Kirchsteiger (1994). His definition of envy corresponds to the current definition of "spite", though.

person context is that indifference curves in (*m, o*) space are positively sloped below and negatively sloped above the 45° line.

Table 1: Archetypes of Distributional Preferences and their Core Properties

preference type	***m* > *o***	***m* < *o***
selfish (own money maximizing)	$\partial u/\partial o = 0$	$\partial u/\partial o = 0$
altruistic (efficiency loving, surplus maximizing)	$\partial u/\partial o > 0$	$\partial u/\partial o > 0$
spiteful (competitive, status seeking, relative income m.)	$\partial u/\partial o < 0$	$\partial u/\partial o < 0$
envious (grudging)	$\partial u/\partial o = 0$	$\partial u/\partial o < 0$
maximin (Rawlsian, Leontief)	$\partial u/\partial o > 0$	$\partial u/\partial o = 0$
inequality averse (inequity averse, egalitarian)	$\partial u/\partial o > 0$	$\partial u/\partial o < 0$
equality averse (equity averse)	$\partial u/\partial o < 0$	$\partial u/\partial o > 0$
kick-down (bully the underlings)	$\partial u/\partial o < 0$	$\partial u/\partial o = 0$
kiss-up (crawl to the bigwigs)	$\partial u/\partial o = 0$	$\partial u/\partial o > 0$

Table 1 lists and Figure 1 displays two further archetypes of distributional preferences, "kick down" and "kiss up". Those types have not been discussed in the literature and are included for completeness only:

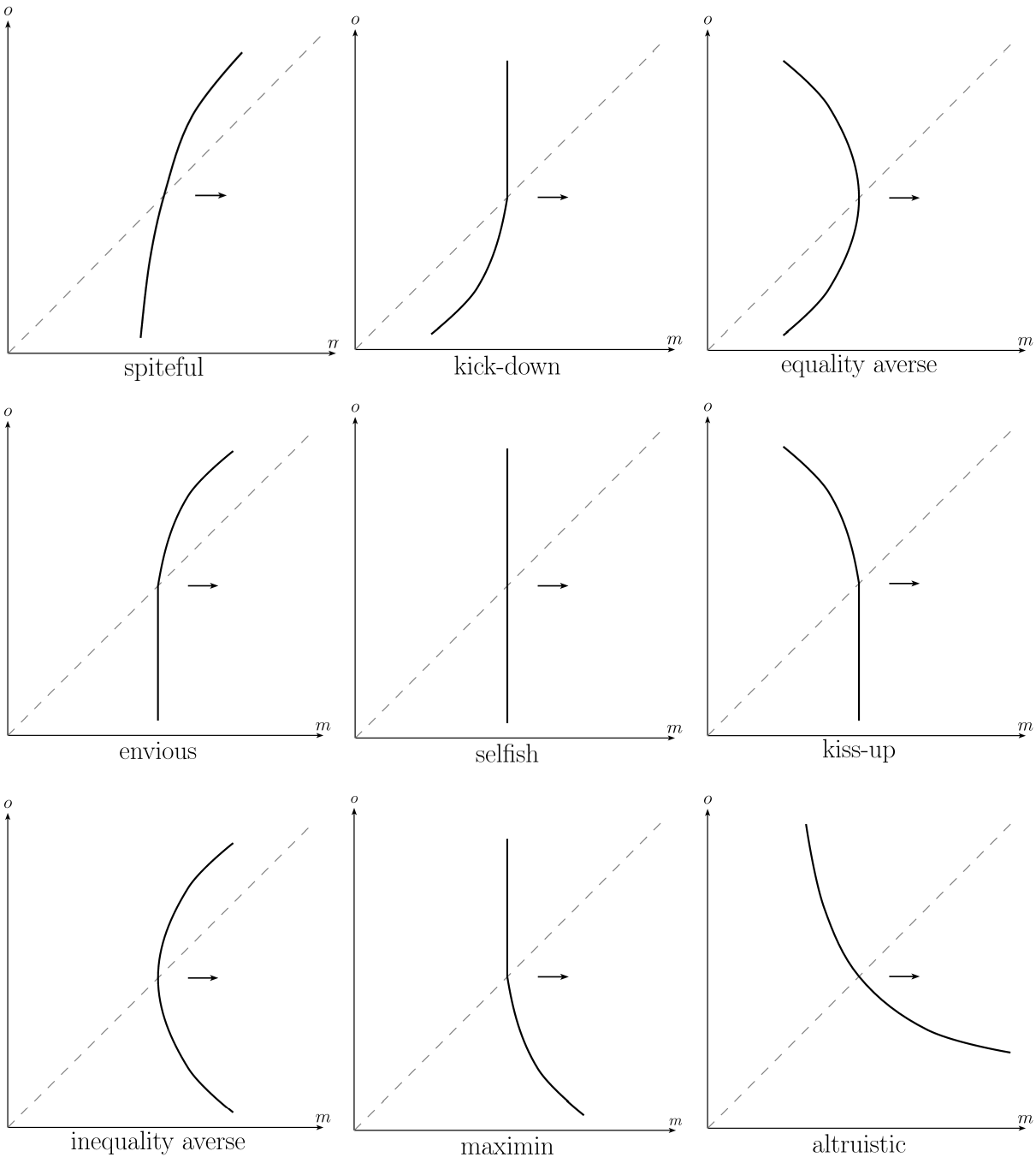
- *Kick-down* or *bully-the-underlings* preferences imply malevolence towards agents who have lower and neutrality towards agents who have higher payoffs. Thus, the defining feature of such preferences in a two-person context is that indifference curves in (*m, o*) space are positively sloped in the domain of advantageous inequality and vertical in the domain of disadvantageous inequality.[15]
- The opposite constellation, benevolence towards agents who are better off combined with neutrality towards those who are worse off, is called *kiss-up* or *crawl-to-the-bigwigs* preferences and such preferences imply negatively sloped indifference curves in the domain of disadvantageous inequality and vertical indifference curves in the domain of advantageous inequality.

Note that the nine types listed in Table 1 and displayed in Figure 1 are well delimitated, mutually exclusive and comprehensive. Also note how the three basic assumptions introduced earlier enter the picture: *ordering and continuity* translate into existence and uniqueness of indifference curves through any point in (*m, o*) space; *strict m-monotonicity* means that upper contour sets are to the right of an indifference curve (the arrows in Figure 1); and *piecewise o-*

[15] A basic disposition related to our "kick-down" preferences has recently been discussed (by Kuziemko et al. 2011, for instance) under the heading "last-place aversion". A last-place averse individual has a psychological disgust against being "last", which creates a propensity for low-income individuals to punish individuals slightly below or above themselves, in the hope of keeping at least one agent below them.

monotonicity requires that the sign of the slope of an indifference curve (i.e., whether the agent is benevolent, neutral or malevolent) changes at most once – when crossing the equal-material-payoff line. Thus, assumptions 1-3 together naturally result in a distinction between the nine mutually exclusive and comprehensive archetypes listed in Table 1 and displayed in Figure 1, meaning that qualitatively there is no room left for additional types.

Figure 1: Typical Indifference Curves of the Nine Archetypes of Distributional Concerns



Arrows → indicate the locus of upper contour sets

There is room left for discussions on names, of course – see Footnote 14 for an example. And there is room left for discrimination within a given class; for instance, it might be interesting

and important to discriminate between altruism and efficiency (or cake-size) concerns (both imply negatively sloped indifference curves in both domains), or between the Fehr and Schmidt (1999) and the Bolton and Ockenfels (2000) model of inequality aversion (both imply positively sloped indifference curves above and negatively sloped indifference curves below the 45° line). Although discrimination within a given class is not the main focus of the paper, Subsection 3.4 discusses a test version that might turn out to be helpful for this task as well.

Before proceeding it seems important to address two potential critiques. One is that the nine archetypes defined above are not really new. This is correct, of course. The main contribution of the present paper is *not* to introduce new preference types; one of the goals is rather to derive the number and core properties of preference types from a small set of primitive assumptions on preferences. This stands in contrast to previous studies which either start with a given list of types or a specific model of preferences. A second – related – critique is that a list of archetypes similar to the one presented in Table 1 could also be obtained by working off the possible sign combinations of the two parameters in the piecewise linear model originally introduced by Fehr and Schmidt (1999) as a description of self-centred inequality aversion and later extended by Charness and Rabin (2002) to allow for other forms of distributional concerns. If one is willing to assume that subjects have preferences of this very specific form then this critique is justified. But, a major point in the current paper is exactly that there is no need to impose such a tight structure.[16] This is true both for the type delineation and for the elicitation procedure. Stated differently, all modelling variants of distributional preferences satisfying the three assumptions introduced in Subsection 2.1 and all distributional archetypes tested for in previous experiments fall into one of the nine categories defined here. This is also true for the Charness and Rabin model, of course. On the other hand, there are many models of distributional preferences in the economic literature that do not fit into the piecewise linear framework of Charness and Rabin – the altruism models by Andreoni and Miller (2002) and Cox and Sadiraj (2007, 2012), the envy model by Bolton (1991), and the inequality aversion model by Bolton and Ockenfels (2000) are prominent examples.

3 Basic Version of the *XY*-Test and Graphical Representation of Test Results

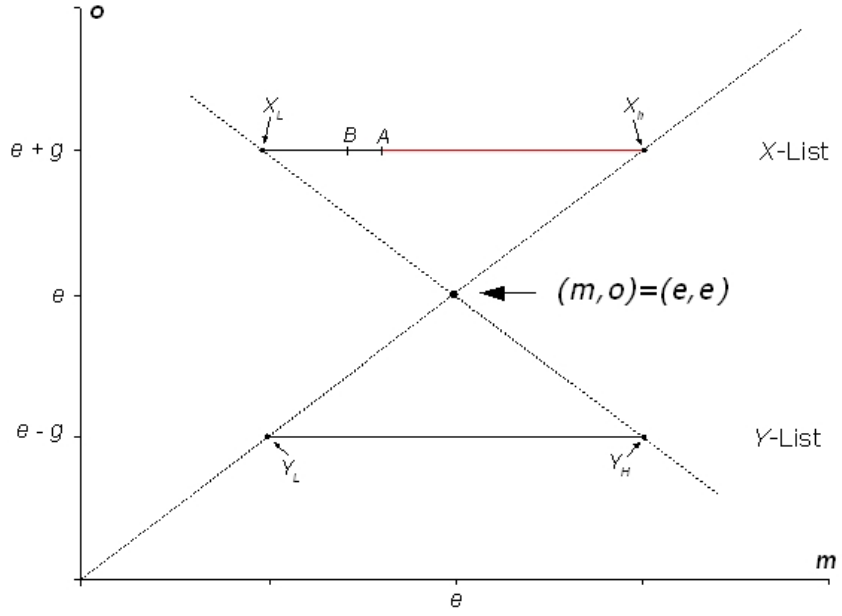
3.1 Idea of the *XY*-Test

As mentioned earlier, the 3 basic assumptions introduced in Subsection 2.1 – the *ordering* part of Assumption 1, *strict m-monotonicity* and *piecewise o-monotonicity* – not only naturally

[16] Using the piecewise linear model in empirical work is not the same as assuming piecewise linear preferences, of course: In experimental work where stakes tend to be small one might argue that the parameter estimates correspond to a piecewise linear approximation of the real preferences. We discuss this point further in Section 5.

result in a mutually exclusive and comprehensive distinction between 9 archetypes of distributional concerns, but also give rise to a clean identification procedure (a "test") that does not rely on unnecessary structural assumptions. This subsection explains how the test works and where the *"XY"* in the name of the test comes from.

Figure 2: Idea of the XY-Test



The aim of the test is to elicit the shape of a *single* indifference curve of the person under scrutiny. To see how it works, take an arbitrary point (*m, o*) = (*e, e*) on the equal-material-payoff line. Then position two parallel lines in the (*m, o*)-space, one above and the other below point (*e, e*), as shown in Figure 2. Suppose the two lines, denoted the *X*-List line and the *Y*-List line, are composed of a large finite number of points. Then the test implicitly asks the decision maker to color red all points on the two lines that she prefers to point (*m, o*) = (*e, e*). *Ordering* makes sure that the decision maker is able to accomplish this task. *Strict m-monotonicity* ensures (i) that the red points on each line – i.e., the elements of the intersection between the upper contour set to the point (*e, e*) and the points on the respective line – are connected to each other and similarly for the black (i.e. not colored) points – the elements of the intersection between the lower contour set and the line; and (ii) that all red points are to the right of the black points. The color of the points on the two lines, or "lists", allows inferring the sign of the slope of the indifference curve in the domain of disadvantageous inequality (*X*-List) and in the domain of advantageous inequality (*Y*-List). *Piecewise o-monotonicity* guarantees that the sign of the slope depends only on the domain and not on the exact location of the two lines. The sign of the slope of the indifference curve in the two domains, in turn, allows distinguishing between the nine archetypes.

Suppose point A (and each point to the right of point A) is colored red, while the point immediately to the left of point A is colored black. Then we can infer from this that in the domain of disadvantageous inequality (that is, above the 45° line) the indifference curve of the DM through point (*e*, *e*) must pass the *X*-List line between point A and the point immediately to the left of A (strict *m*-monotonicity ensures that there is at most one point on the *X*-List line where the DM is exactly indifferent to the equal-material-payoff allocation under consideration).[17] This, in turn, implies that this DM is either altruistic, or equality averse, or kiss-up (because those are the only archetypes that imply negatively sloped indifference curves above the 45° line – see Figure 1). The other six possibilities listed in Table 1 and displayed in Figure 1 are inconsistent with her behavior. Adding to this the DM's choices in the domain of advantageous inequality (the *Y*-List) discriminates between the remaining three possibilities.[18] Basically, the behavior of the DM on the *X*-List (representing the domain of disadvantageous inequality) tells us in which column in Figure 1 her type is in, while her behavior on the *Y*-List tells us the row. Also note that the switch points on the two lists reveal not only the archetype of distributional preferences but also give information on preference intensities. For instance, an individual who switches between black and red on point B is more benevolent in the domain of disadvantageous inequality than an individual who switches on point A. As will be shown in Subsection 3.3, this information can be used to construct a two-dimensional index representing both, archetype and preference intensity.

3.2 The Symmetric Basic Version of the *XY*-Test

Although feasible in principle, asking subjects to color a subset of points (those preferred to the given equal-monetary-payoff allocation) out of a larger set (the *X*-List or the *Y*-List) might be too demanding a task in a lab experiment. Also, the choice sets themselves (in Figure 2, the location of the points X_L, X_H, Y_L and Y_H) might shape behavior. The proposed test therefore exposes subjects to a series of diagnostic binary choice problems. In the (symmetric) basic version of the test the family of binary choices is characterized by four positive integers, *e*, *g*, *s* and *t*, where

(i) *e* determines the locus of the *equal-material-payoff allocation* (*m*, *o*) = (*e*, *e*);

[17] If the whole *X*-List line is red, we can infer from this that the upper contour set in the domain of disadvantageous inequality extends at least to point X_L, if the whole *X*-List line is black then we can infer that the lower counter set in this domain extends at least to point X_H.

[18] The only case where some ambiguity is left on the exact sign of the slope of the indifference curve in the domain of disadvantageous inequality (advantageous inequality, respectively) is when the border between black and red points is exactly above (below, respectively) the point (*m*, *o*) = (*e*, *e*). Below a DM is referred to as *weakly benevolent* (malevolent) *in the domain of disadvantageous inequality* if she decides such that (i) the border between black and red points is exactly above the point (*m*, *o*) = (*e*, *e*) and (ii) the point exactly above (*m*, *o*) = (*e*, *e*) is colored red (black, respectively). A similar convention is used for the domain of advantageous inequality (where a red point exactly below the reference point is associated with malevolence and black with benevolence). Here note that a weakly benevolent DM reveals benevolence only in the impartial decision where no own money is at stake. Also note that without explicitly asking for indifferences, vertical indifference curves can never be identified for sure; but they can be identified with arbitrary precision – see the discussion in the next subsection.

(ii) g is a *"gap" variable* characterizing the vertical distance between (e, e) and the two lists – see Figure 3; in order to avoid zero or negative monetary payoffs it seems sensible to restrict g to be strictly smaller than e;

(iii) s is a *"step size" variable* characterizing the horizontal distance between two adjacent points on a list (a restriction on s is imposed in the next point);

(iv) $t \geq 1$ is a *"test size" variable* determining the number of steps (of size s) which are made to the left and to the right starting from the point just above or below (m, o) = (e, e); in order to preserve advantageous and disadvantageous inequality it seems sensible to impose the restriction $t \leq g/s$.

In total the test consists of $4t + 2$ binary decision problems. In each decision problem the subject is asked to decide between two alternatives (named Left and Right), each involving a payoff pair – one payoff for the subject (the DM) and one for the (randomly matched, anonymous) other subject (the passive person). For expositional purposes the decision problems are separated into two blocks, the disadvantageous inequality block (*X*-List) and the advantageous inequality block (*Y*-List). Within each block the decision problems are presented as rows in a table. In each decision problem one of the two alternatives (the alternative "Right", say) is the (recurring) equal-material-payoff allocation (m, o) = (e, e). The second alternative in each decision problem (the alternative "Left") is constructed as shown in tables 2 and 3.

Figure 3: The Geometry of the XY-Test

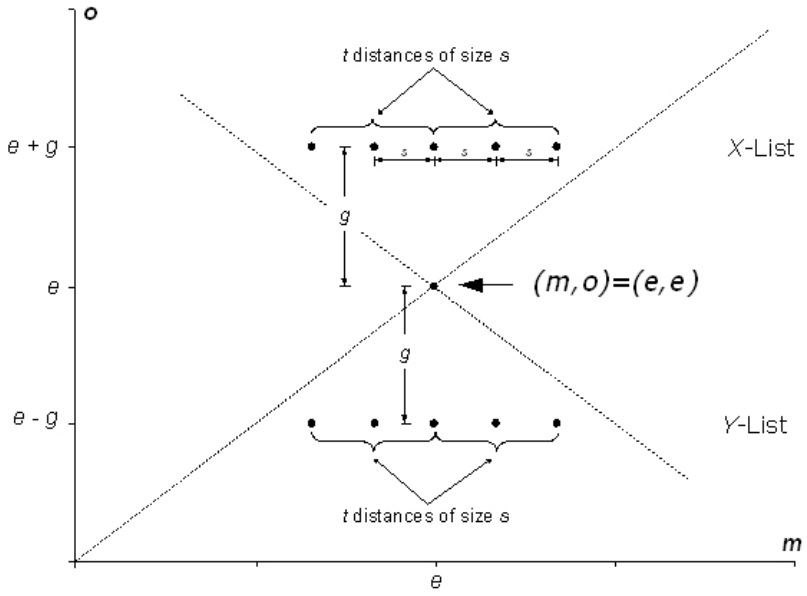


Table 2: The *X*-List (Disadvantageous Inequality)

Alternative: **Left**			Alternative: **Right**		
please mark below if you prefer Left	you receive tokens	the passive person receives tokens	you receive tokens	the passive person receives tokens	please mark below if you prefer Right
□	$e - ts$	$e + g$	e	e	□
…	…	…	…	…	…
□	$e - s$	$e + g$	e	e	□
□	e	$e + g$	e	e	□
□	$e + s$	$e + g$	e	e	□
…	…	…	…	…	…
□	$e + ts$	$e + g$	e	e	□

Table 3: The *Y*-List (Advantageous Inequality)

Alternative: **Left**			Alternative: **Right**		
please mark below if you prefer Left	you receive tokens	the passive person receives tokens	you receive tokens	the passive person receives tokens	please mark below if you prefer Right
□	$e - ts$	$e - g$	e	e	□
…	…	…	…	…	…
□	$e - s$	$e - g$	e	e	□
□	e	$e - g$	e	e	□
□	$e + s$	$e - g$	e	e	□
…	…	…	…	…	…
□	$e + ts$	$e - g$	e	e	□

An important feature of the test is that within each of the two blocks the material payoff of the passive person in the asymmetric allocation is held constant, while the material payoff of the DM increases monotonically from one choice to the next. Together with the fact that the symmetric allocation remains the same in all choices, this design feature guarantees that strict *m*-monotonicity is enough to make sure that when facing the choice between Left and Right within a given block, each individual switches at most once from Right to Left (and never in the other direction). Here note that deciding for Left in a given row of Table 2 (Table 3, respectively) is equivalent to coloring red the corresponding point on the X-List line (Y-List line) in Figure 3.

The minimum test size has $t = 1$ yielding 6 binary decision problems. If classifying subjects into one of the nine archetypes is the main aim of a study then there is no need to use a larger test design since observing the behavior of a DM in 6 binary choices already allows *discriminating between* the nine *archetypes at* any *arbitrary precision*. More specifically, the researcher needs to define when an agent should be considered as egoistic in a particular domain (this is the meaning of arbitrary precision). Suppose we define an agent to be egoistic in a particular domain if she is not willing to give up c Cents in order to change the material payoff of the passive person by 1$. Then the appropriate 6 items test has to be such that $c =$

$100s/g \Leftrightarrow s = cg/100$ meaning that we can choose two of the remaining three parameters (after having fixed t at 1) freely.

3.3 Identifying Archetype and Characterizing Intensity of Distributional Concerns: The (x, y)-Score

This subsection describes a method to identify archetype and to characterize intensity of distributional preferences at the individual level and a procedure to represent the type-intensity distribution graphically.

Step 1 (Consistency Check): As argued above an individual whose preferences satisfy strict *m*-monotonicity has at most one switch from Right to Left (and no switch in the other direction) in each of the two tables. Step 1 is to eliminate all subjects that fail this basic consistency check (in an implementation of the symmetric basic version of the test – see Section 6 for details – less than 5% of the subjects failed the consistency check).

Table 4: Determination of (x, y)-Score

subject chooses Left for the 1st time in row	in the *X*-list (*x*-score)	in the *Y*-list (*y*-score)
1	$t + 0.5$	$-(t + 0.5)$
2	$t - 0.5$	$-(t - 0.5)$
...	...	...
t	1.5	- 1.5
$t + 1$	0.5	- 0.5
$t + 2$	- 0.5	0.5
...	...	...
$2t + 1$	$-(t - 0.5)$	$t - 0.5$
never	$-(t + 0.5)$	$t + 0.5$

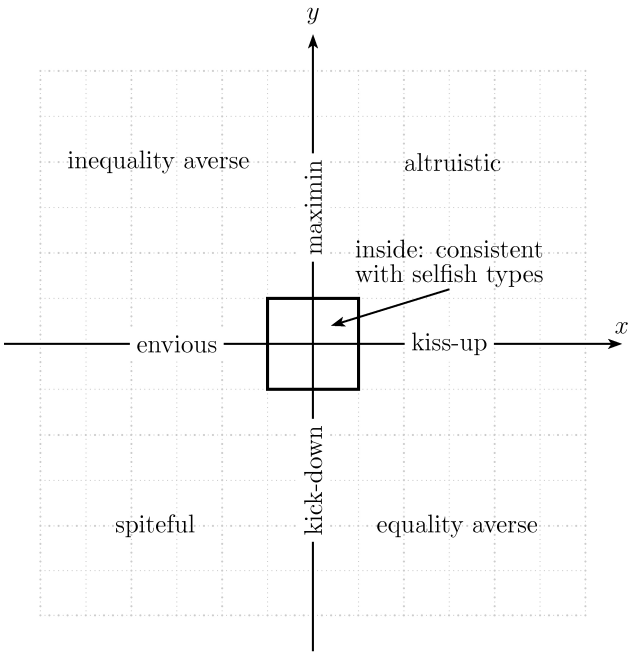
Step 2 (Defining Scores): Represent each subject with consistent behavior by an (x, y) tuple defined as follows: The *variable x* (*x-score*) summarizes the behavior of the individual in the disadvantageous-inequality related block (*X*-List) and is defined as (t + 1.5) points minus the row number in which the individual decides for the first time for the asymmetric allocation (that is, for the payoff vector on the left hand side). If an individual always decides for the symmetric (or egalitarian) allocation, we take the convention that she decides for the first time for the asymmetric allocation in the $(2t + 2)^{th}$ row, so that she gets an *x*-score of $-(t + 0.5)$. For instance, if in the test version displayed in Figure 3 (where $t = 2$) an individual decides for the symmetric allocation in the first row of the *X*-List and for the asymmetric allocation in the second (and in all other) row(s) then she gets an *x*-score of 3.5 – 2 = 1.5. The *variable y* (*y-score*) summarizes the behavior of the subject in the advantageous-inequality related block (the *Y*-List) and is defined as the row number in which the individual decides for the first time for

the asymmetric allocation minus (t + 1.5) points. If an individual always decides for the symmetric allocation, we take again the convention that she decides for the first time for the asymmetric allocation in the $(2t + 2)^{th}$ row; she then gets a y-score of t + 0.5.

Note that the definition of the two scores implies that each of them can take on $2(t + 1)$ different values (see Table 4); thus, the proposed test allows for $4(t + 1)^2$ different (x, y)-scores. Also note that the sign of the x-score corresponds to the sign of $\partial u/\partial o$ in the domain of disadvantageous inequality, while the sign of the y-score corresponds to the sign of $\partial u/\partial o$ in the domain of advantageous inequality. Furthermore, the magnitude of the x-score (y-score, respectively) is an ordinal index of the intensity of distributional preferences in the domain of disadvantageous inequality (advantageous inequality, respectively).[19]

Step 3 (Representing Relative Frequencies of Types): Represent the absolute or relative frequencies of the different (x, y)-scores in an axis of abscissas as shown in Figure 4.

Figure 4: Distributional Types in (x, y) Space



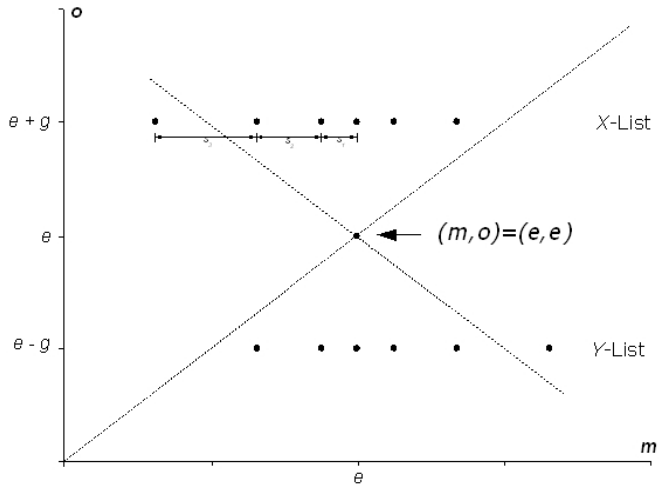
3.4 Extending and Refining the Test

This subsection proposes three modifications of the symmetric basic version of the test that might help to shed light on more specific research questions. The first modification replaces the symmetric step-size in the basic version by an asymmetric one, the second modification extends the X-List to the left and the Y-List to the right and the third modification adds more lists.

[19] Also note that a score of +0.5 indicates weak benevolence, while a score of -0.5 indicates weak malevolence.

In the (symmetric) basic version of the test, the DM's trade-off between her own and the passive person's material payoff depends on the relation between the horizontal distance between two neighboring dots in Figure 3 (“step size”, *s*) and the vertical distance between those dots and the equal-material-payoff point ("gap size", *g*). By increasing *g*, keeping the rest of the test as it is (or by decreasing *s*, keeping the rest as it is) the power of the test to discriminate between selfish and different variants of non-selfish behavior is increased (remember the discussion on "identification with arbitrary precision" in Subsection 3.2). However, if the test size is held constant this comes at a cost since the discriminatory power of the test at the borders is decreased (too many subjects end up with extreme values of *x*- or *y*-score). Here an *asymmetric step-size version* of the test – where step size is small at the centre but grows larger when moving away from the centre – might be a good compromise. See Figure 5 for an illustration.

Figure 5: Asymmetric Step- and Test-Size

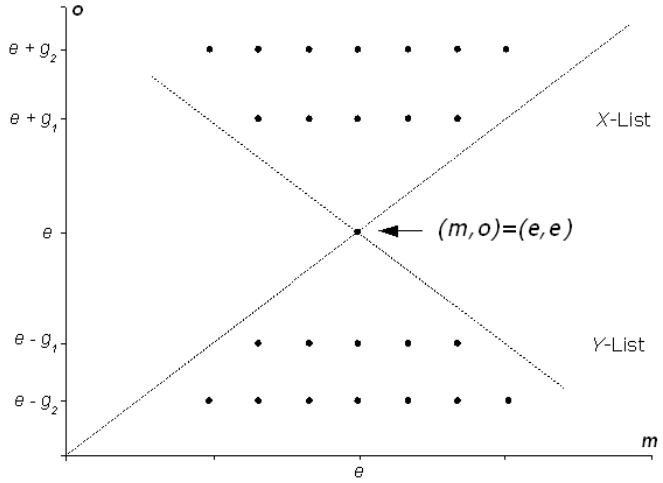


For some research questions it might be interesting to know whether there are subjects who (in the relevant range) put more weight on the material payoff of the passive person than on their own material payoff (that is, whether there are subjects for whom $\partial u/\partial o > \partial u/\partial m$). An *asymmetric test-size version* of the procedure, where the *X*-List is extended upward – so that in (*m*, *o*) space the alternatives considered in the test cross the -1 line – while the *Y*-List is extended downwards – so that in (*m*, *o*) space the alternatives considered in the test cross the -1 line – might help to answer this question.[20] More specifically, consider the following version of the test: The variables *e*, *g* and *s* are as in the (symmetric) basic version of the test, *t* (the "test-size variable) is no longer a choice variable but is taken as the largest integer smaller than *g/s*, and the new variable *a* (the "asymmetry variable") indicates how many choices are added on

[20] Here note that lifting the restriction $ts < g$ and extending both lists symmetrically elongates the *X*-List (*Y*-List, respectively) into the domain of advantageous (disadvantageous) inequality which is not what we want to have.

each list (starting from the symmetric basic version of the test). The test now consists of a series of $4t + 2a + 2$ decision problems divided into two blocks (*X*-List and *Y*-List), each containing $2t + a + 1$ pairs of alternatives. The alternatives are exactly as displayed in tables 2 and 3 with the exception that the *X*-List starts with $m = e - ts - as$ (as before, it ends with $m = e + ts$) and that the *Y*-List ends with $m = e + ts + as$ (as before, it starts with $m = e - ts$). An important feature of the asymmetric test-size version is that it contains decision problems for which strong altruism dictates a different choice than strong surplus-maximization concerns. For instance, in the first decision problem on the *X*-List of the test version displayed in Figure 5, a surplus maximizer has to choose the point (*e, e*) (because this option maximizes both *m* and *m*+*o*), while a strong altruist might be inclined to color red the leftmost point on the *X*-List (because this point maximizes *o*). A similar argument applies to the rightmost point on the *Y*-List. Thus, the asymmetric test-size version can potentially help to shed light on the question whether those who are willing to help others are driven by surplus-maximization motives (Charness and Rabin 2002's 'social-welfare preferences', or Engelmann and Strobel 2004's 'taste for efficiency') or by pure altruism (Becker 1974, Andreoni and Miller 2002).

Figure 6: A Multi-List Variant of the XY-Test



For some purposes (for instance, to discriminate between different parametric forms within a given class, as, e.g. between the Fehr and Schmidt 1999, and the Bolton and Ockenfels 2000 model of inequality aversion – both imply indifference curves that have negative slope below and positive slope above the 45° line) it might be interesting to gain more insights on the exact shape of indifference curves in *(m, o)*-space. With a *multi-list version* of the test, where subjects

are asked to complete two or more *X*- and *Y*-lists distinguished by the size of the gap variable *g* (see Figure 6 for an example), indifference curves can be elicited with arbitrary precision.[21]

4 Discussion of the Test in View of the Literature

The current paper contributes to the growing literature on identifying distributional preference types in the lab. This section highlights the main differences between the current approach and other approaches proposed in the literature.

4.1 Aggregate Level Studies in Search for "the Universal Utility Function"

Charness and Rabin (2002), Engelmann and Strobel (2004), Cox, Friedman and Gjerstad (2007) and Cox and Sadiraj (2012) search for utility functions that are good descriptions and predictors for their aggregate data and they find that, on average, negatively sloped indifference curves (which correspond to entries in the positive quadrant in Figures 4) predict behavior better than any other constellation. While this is an interesting and important insight per se, the issue of preference heterogeneity is left unaddressed by those studies. In this respect the non-parametric identification approach introduced here is complementary to this literature. If one searches for a unifying model that describes aggregate behavior fairly well, one of the models proposed in this literature is a good choice.[22] For those more interested in the heterogeneity of preferences and behavior an identification approach – as the one proposed here – seems more useful. It has the advantage that subjects who do not fit the average preference type (most importantly, inequality averse, envious, and spiteful agents) can be identified and their behavior can correctly be interpreted. This is important because – depending on the design of the game – the preferences and behavior of even a small fraction of agents might have a large impact on the outcome. For instance, as experimental economists have successfully shown, a small

[21] Of course, with several *X*- and *Y*-lists inconsistent choices (more than one switch in one of the lists or switches in the wrong direction) might become more important than they are in the basic version and econometric techniques (such as finite mixture models) might be needed to get meaningful results. We discuss this point further in Section 5.

[22] Charness and Rabin (2002) propose a "*quasi-maximin model*" in which utility increases in the lowest of all agents' monetary payoff (the maximin property) and the total of all agents' payoffs (the surplus maximization property). Engelmann and Strobel (2004) present evidence that supports the view that the behavioral impact of both efficiency and maximin concerns is stronger than that of inequality aversion in the aggregate. Cox and Sadiraj (2012) present experimental evidence that is inconsistent with inequality aversion and quasi-maximin models and use this evidence to motivate their "*egocentric altruism model*" in which preferences are positively monotonic (i.e., utility is increasing in the material payoffs of all agents), strictly convex (utility is strictly quasi-concave in all its arguments), and "egocentric" (if a player has the choice between two *(m,o)* allocations where the *m* part in one allocation is equal to the *o* part in the other and vice versa, then he chooses the one with the higher *m*). Cox, Friedman and Gjerstad (2007) propose a similar CES utility function but allow an agent's distributional preferences to depend on relative status, on previous behavior of others (reciprocity) and on the set of alternative actions available to others.

fraction of inequality averse players in a public good game with punishment is sufficient to credibly threaten that free riders will be punished, inducing even selfish agents to contribute to the public good (see Fehr and Gächter 2000b and 2002, or Fehr and Fischbacher 2003, for instance). Similarly, as shown by Tyran and Sausgruber (2004) "a little fairness" (i.e., a small fraction of inequality averse agents) may induce a lot of redistribution in democracy. Also, as shown by Malmendier and Szeidl (2008) a small fraction of spiteful agents can have a large impact on the outcome of an auction because the auction format "fishes for fools" (i.e., disproportionately many buyers with an overbidding bias end up as winners). In such circumstances a utility or motivational function that fits well on average is not likely to be a good predictor for the outcome of the game (because the preferences of a minority have a disproportional impact on the outcome).

4.2 Continuous Dictator Game Studies Looking at Heterogeneity at the Individual Level

Andreoni and Miller (2002) expose their subjects to a series of 8 dictator games differing in budget size and price of giving. Their main aim is to test whether subjects' behavior is consistent with the General Axiom of Revealed Preference (GARP) and one of their remarkable results is that this is indeed the case with most subjects. Furthermore, they investigate preference heterogeneity at the individual level. Making only mild assumptions on how to look at the data they find that the behavior of nearly half of the subjects fits a textbook CES utility function without error. Specifically, the behavior of 14% of the subjects is perfectly consistent with maximization of the 'Leontief' functional form $u(m, o) = min\{m, o\}$, the behavior of 6% of the subjects is consistent with the 'perfect substitutes' form $u(m, o) = m + o$, and the behavior of 23% of the subjects is consistent with the 'selfish' form $u(m, o) = m$. The remaining subjects are then classified as weak incarnations of one of those three pure types according to the Euclidian distance between actual behavior and behavior predicted by the pure types. This is an interesting and important study with a slightly different focus, though: While Andreoni and Miller's main focus is on costly giving, our interest is more symmetric on both, costly giving and costly taking. This manifests itself in the results. For instance, while their design is perfectly suited to discriminate between different varieties and intensities of altruistic preferences, it is less well suited to distinguish between selfish, spiteful and envious types because those types always contribute zero in a costly-giving game.[23] Also, as argued by List

[23] To get additional insights, Andreoni and Miller also expose subjects to upward sloping budget lines. However, since indifference curves of spiteful and envious agents are upward sloping as well and since they are not necessarily convex there is no guarantee that such a design correctly classifies agents. It is noteworthy, though,

(2007) and others, their dictator game design might produce experimenter demand effects because asking a subject to give money to another person if she wants to might make the subject believe that the experimenter expects her to do something – and because the only thing she can do is to share a positive amount of money, she does share money, even if she is selfish or spiteful.

Fisman, Kariv, and Markovits (2007) extend Andreoni and Miller (2002)'s analysis by *(i)* letting subjects play many more dictator games (each subject is confronted with 50 independent decision problems); *(ii)* also considering three-person dictator games; and (iii) allowing for non-linear (step-shaped) budget sets. The latter extension allows them to classify subjects into more distributional types (they distinguish between 'self-interested': 13% of their population; 'lexself': 49%; 'social welfare': 13%; 'competitive': 6%; and 'mixed preferences': 19%). While this is clearly an important study that sheds new light on a variety of interesting aspects, the classification procedure employed requires a substantial amount of time and effort and therefore seems less suited to be used as a tool in experimental economics to control for subject pool effects, or to help to interpret data from other (unrelated) experiments. Also, since there are no upward sloping parts in their budget sets it is impossible to observe choices where a subject gives up own monetary payoff to reduce the monetary payoff of the other player; thus, competitive, envious or even inequality averse types cannot be cleanly identified with this design.[24]

4.3 Discrete Choice Studies Looking at Heterogeneity at the Individual Level

Blanco, Engelmann, and Normann (2011) ran four different experiments (ultimatum game, modified dictator game, sequential prisoner's dilemma game and public goods game) with the same sample of subjects with the aim of testing the Fehr and Schmidt (1999) piecewise linear model of inequality aversion. Their main finding is that the Fehr and Schmidt model has considerable predictive power at the aggregate level but performs less well at the individual level. While this is an important insight *per se* it does not give a definitive answer to the question whether preferences are unstable on the individual level or simply do not conform to the tight structural assumptions imposed by the Fehr and Schmidt model (in addition to the piecewise linear form, it does not allow for parameter estimates that fall outside the North/West

---

that the empirical results reported in Section 6 are roughly in line with the Andreoni and Miller results, indicating that their design does quite a good job.

[24] Their design has other virtues. For instance, by exposing DMs to a large number of dictator game tasks it allows investigating consistence with GARP (the authors find that most subjects' choices nearly satisfy GARP), a task for which the test proposed here is neither suited nor intended to be suited. Also, in one of their treatments DMs are exposed to choices that have consequences for three persons, for the DM and for two other subjects. This allows them to investigate the relationship between 'preferences for giving' (tradeoffs between the DM's own payoff and the payoffs of others) and 'social preferences' (trade-offs between the payoffs of others).

quadrant of Figure 4). Also, while the parameter of aversion to advantageous inequality is estimated by Blanco et al. in an environment uncontaminated by intentions and beliefs (similar to the approach proposed here),[25] the parameter of aversion to disadvantageous inequality is estimated from second-mover behavior in a sequential-move strategic game – the ultimatum game – where the imputation of intentions behind the action of the first mover or the interpretation that certain actions are likely to come from certain types can affect the behavior of the second mover.[26]

Iriberri and Rey-Biel (2013) present a discrete choice test intended to classify individuals into selfish, social welfare maximizers, inequality averse and competitive. For this purpose they expose subjects to 16 choices among three *(m, o)* allocations. In terms of our Figure 3 the three *(m, o)* allocations in each decision task are located in a ":·" pattern to each other. That is, two of the three allocations have the same *m* ($m_1 = m_2 = m^*$, say*)* but a different *o* ($o_1 = o^* + \lambda$ and $o_2 = o^* - \lambda$ with $\lambda > 0$, say), the third option adds 1 ECU to m^* and averages the *o* of the other two options (that is, $m_3 = m^* + 1$ and $o_3 = o^*$). The 16 decision tasks differ in the location of m^* and o^* and in the parameter λ. The authors then use three different mixture-of-types models to estimate distributional preference parameters under the assumption that subjects decide according to Charness and Rabin's (2002) piecewise linear motivational function. In the individual by individual estimation they identify 44% of the population as selfish, 21% as social welfare maximizers, 25% as inequality averse and 10% as competitive. This is an interesting study yielding important insights about heterogeneity of distributional preferences. A salient difference in the identification procedures is that the binary decision tasks proposed here are ordered and designed to elicit the shape of a *single* indifference curve in the (*m, o*)-space while their trinity decision tasks are distributed all over the (*m, o*)-space and

[25] Their estimate of the parameter of aversion against advantageous inequality is theoretically elegant and related to the approach used here: In terms of Figure 3, Blanco et al.'s 20 binary decisions correspond to 20 points on the 45° line through the positive orthant each compared to a single point that has exactly the same *m* as the rightmost point on the 45° line but an *o* of zero. Thus, by observing the choices of a subject in the 20 binary decision problems one can deduce the shape of the indifference curve through this latter point. While this is an elegant procedure it has some disadvantages in comparison to the approach proposed here, namely (i) that the shape of the indifference curve can only be assessed for the domain of advantageous inequality; and (ii) that positive sloped indifference curves in that domain cannot be identified correctly. Blanco et al. indeed find that roughly 10% of the subjects prefer the point on the axis of abscissas to the point vertically above that point, a behavior consistent with positive sloped indifference curves in that domain. Note that this figure is remarkably close to the one we obtain in the experiments reported in Section 6 (roughly 10% of the subjects have a negative *y*-score - see Figure 7).

[26] This should not be read as a critique against their approach. After all the Fehr and Schmidt model has been designed to explain behavior in strategic games and the "calibration" of parameters with ultimatum game data has been suggested by the inventors themselves. A possible advantage of their approach (in terms of explaining behavior in strategic games) is that it captures not only distributional preferences but also other forms of other-regarding preferences (such as reciprocity motives, for instance).

designed to yield data for the estimation of the parameters of a given model.[27] A related difference is that their identification procedure relies on tight structural assumptions, while the non-parametric approach proposed here does not rely on *any* structural assumptions.[28]

The ring-test has been developed by social psychologists (standard references are Griesinger and Livingston 1973 and Liebrand 1984) to assess "social value orientations" at the individual level, and it has been employed by economists in a variety of different experiments to identify type and measure intensity of distributional preferences (see, for instance, Offerman et al. 1996, Sonnemans et al. 1998, van Dijk et al. 2002, Brosig 2002, Brandts et al. 2009, or Sutter et al. 2010). In its simplest implementation (often called the "circle-test") the ring-test asks subjects to choose their most preferred point on a circle in the (*m, o*) space. The circle has its centre at the origin of the (*m, o*) plane and has a radius of *r*, say. Depending on their choice subjects are then classified either into the 5 "*social value orientation"* (*SVO*) *types*: "altruistic", "cooperative", "individualistic", "competitive" and "aggressive", or into 8 SVO types, the 5 listed types plus "martyrdom", "masochism" and "sadomasochism". Although interesting from an aesthetical point of view, the ring-test has serious theoretical shortcomings, the most important one being that the test assumes linear preferences (see Appendix B for details) and thereby systematically misclassifies those archetypes that are inconsistent with linearity (e.g., because they imply a sign change in $\partial u/\partial o$). For instance, inequality averse subjects are typically classified by the ring-test – depending on the intensity of preferences – either as altruistic ("cooperative" in the language of the ring-test) or as selfish ("individualistic" in the language of the ring test), while inequality loving subjects are classified either as competitive (here the ring-test language corresponds to our language) or as selfish. Furthermore, the ring-test cannot distinguish between maximin and egoistic preferences (both are classified as "individualistic" types), or between envious and competitive preferences (both are classified as competitive types). Finally, archetypes that are consistent with linearity, in principle, might be misclassified if observed choices are only consistent with nonlinear indifference curves (as it is

[27] The same is true for the test used by Cabrales et al. (2010). There, subjects are exposed to 24 choices among four (*m, o*) allocations drawn at random (but not uniformly) from a subset of the positive orthant, and evidence from these choices is then used to estimate subjects' distributional preference parameters within the realm of Charness and Rabin's piecewise linear basic model (without reciprocity).

[28] The econometric approaches employed by Cabrales et al. (2010) and Iriberri and Rey-Biel (2013) have other virtues. For instance, by exposing subjects to a large number of tasks they allow for a formal statistical framework within which the degree of consistency of observed choices with preferences represented by a given functional form can be tested. By contrast, the main strength of the approach proposed here is to start with a small set of basic assumptions on preferences and to devise a procedure that assigns archetypes to people in a non-parametric way, given their preferences satisfy the assumptions. That is, our procedure does not –and is not intended to– test whether subjects' preferences satisfy those assumptions (the multiple-list version of the test proposed in Subsection 3.4. would allow for such tests, though). In that sense our approach is similar in spirit to the ones proposed by Holt and Laury (2002) and Dohmen et al. (2010, 2011) for the domain of risk preferences, while the approaches of Cabrales et al. (2010) and Iriberri and Rei-Biel (2013) are more similar to the one used by Bruhin et al. (2010).

the case with convex altruism, for instance).[29] How serious those problems are depends on the context, of course. After all an inequality averse subject, for instance, behaves as an altruistic one in the domain of advantageous inequality but it behaves as a competitive one in the domain of disadvantageous inequality. The same is true of other inappropriately classified types.[30]

Summing up, many interesting approaches to identify archetypes and to measure intensity of distributional preferences have been proposed in the literature but most of them require more time and effort from a subject to produce a score and therefore seem less suited as a tool in experimental economics to control for subject pool effects, or to help to interpret data from other (unrelated) experiments. Furthermore, the identification procedures typically discriminate between a somewhat arbitrary set of preference types or they rely on by far more demanding assumptions on preferences than the approach proposed here.

5 Relation of (*x*, *y*) Score to Other Measures of Type and Intensity of Distributional Concerns

5.1 The *'More Altruistic Than'* Relation by Cox et al. (2008)

In their 2008 paper Cox, Friedman and Sadiraj formalize the binary relation "*more altruistic than*" between two different preference orderings over income allocation vectors. For a given domain $D \subset R^2_+$ this binary relation induces a partial ordering on admissible preferences over (*m, o*) tuples. For the case of negatively sloped indifference curves (altruism, taste for efficiency), preference ordering *A* is *more altruistic than* preference ordering *B* if *A* has shallower indifference curves than *B* in (*m, o*)-space, so *A* indicates a higher willingness to give up units of *m* for a unit increase in *o* than does *B.* Similar for the other eight archetypes of distributional preferences (for the case of positively sloped indifference curves the binary relation *more altruistic than* translates to "less malevolent than"). As is easily verified, if *D* is

[29] Misclassifications are especially likely in the most widely used implementation of the test where subjects are not asked to choose the most preferred point on the circle but are rather exposed to a series of binary choices between adjacent (*m, o*) alternatives on the circle. In this case the standard identification procedure heavily uses the fact that with linear indifference curves the most preferred point on the circle lies exactly opposite to the least preferred one (see Appendix B for a discussion), a property that is typically violated if indifference curves are nonlinear.

[30] Besides the conceptual problems inherent in the test itself, there are also some issues in the typical implementation that seem problematic. For instance, many researchers (i) use the double role assignment protocol, (ii) pay out each single decision and (iii) use fixed pairings throughout the whole classification procedure. Thus, the payoff received by a subject is determined by *all* the decisions she makes in the test and by *all* the decisions made by her "partner". There are at least two potential problems with this, first paying out all decisions means that in each binary decision a subject might not decide for the most preferred allocation but might want to implement her most preferred final allocation; and secondly, in the fixed pairings design not only preferences but also beliefs play a role in the decisions (suppose, for instance, a subject is inequality averse; if she expects the partner to behave selfishly, she will decide selfish too in order to implement a fair overall allocation; by contrast, if she expects the partner to act altruistically, she might have an incentive to act altruistically too, since altruistic acts then lead to a more egalitarian overall allocation).

the entire positive orthant then a necessary condition for the preference ordering of DM i to be *more altruistic than* the preference ordering of DM $j \neq i$ is that $x_i \geq x_j$ and $y_i \geq y_j$.[31]

5.2 Parameter Ranges in Piecewise Linear Model and Willingness to Pay

The (x, y)-score as defined in Subsection 3.3 is an ordinal index of preference intensity (a higher x means a higher weight on the other's payoff when the DM is behind while a higher y means a higher weight on the other's payoff when the DM is ahead) and as such is not normalized with respect to the four design parameters (e, g, s, t). This makes it difficult to compare the results of studies which use different sets of design parameters. This might be regarded as a drawback as the proposed test design is per se well suited for measuring the distributional preferences in experiments and representative surveys with large samples. To make the results of different studies comparable (even if they use different sets of design parameters) it might be advisable to replace the (x, y)-score by a cardinal metric that is equally easy to compute and has a similar intuitive interpretation. One way to get to such a metric is to translate the (x, y)-score into parameter ranges in structured models frequently used in the literature. The most widely used functional form in the empirical literature (see, for instance, Cabrales et al. 2010, Blanco et al. 2011 and Iriberri and Rey-Biel 2013) is the piecewise linear model introduced by Fehr and Schmidt (1999) as a description of self-centered inequality aversion and extended by Charness and Rabin (2002) to allow for other forms of distributional concerns. In the reciprocity free version the Charness and Rabin (CR) representation of preferences takes the form

$$u_{\gamma,\sigma}(m, o) = (1 - \sigma I_{m \leq o} - \gamma I_{m > o})m + (\sigma I_{m \leq o} + \gamma I_{m > o})o, \qquad (CR)$$

where γ and σ are parameters assumed to satisfy $\sigma \leq \gamma < 1$ and where I is an indicator variable that takes the value of one if the condition in the subscript is met and the value of zero otherwise. This formulation says that the DM's utility is a linear combination of her own material payoff and the other person's material payoff and that the (otherwise constant) weight the DM puts on the other's payoff might depend on whether the other is ahead or behind. If one is willing to assume that subjects' preferences can be approximated by this form, how do (x, y)-scores translate into parameter ranges in this model? This question is easily answered. Consider the X-List first. In this domain a DM with CR-preferences weakly prefers LEFT to RIGHT in row $r \in \{1,..., 2t+1\}$ iff $(1 - \sigma)[e + (r\text{-}t\text{-}1)s] + \sigma(e + g) \geq e$. Thus, assuming that a DM who is

[31] For restricted domains of preferences (for instance, if only piecewise linear utility functions are admitted as in Fehr and Schmidt 1999 or in Charness and Rabin 2002, or if only specific kinds of CES functions are considered as in Andreoni and Miller 2002, in Fisman et al. 2007, or in Cox and Sadiraj 2007) combined with a test design with high resolution (large gap variable g combined with small step size s) and for restricted domains of income allocation vectors (adapted to the binary decisions in the test) the two notions are equivalent to each other.

indifferent decides for LEFT, the relationship between x-score and parameter range of σ in the piecewise linear model is as shown in Table 5. Using the same tie breaking rule (an indifferent DM decides for LEFT) for the Y-List we get a similar table (not shown) with x-score replaced by y-score, σ replaced by γ, and strict inequalities replaced by weak ones (and vice versa).

Table 5: x-Score, Parameter σ and Willingness to Pay (WTP^d) in Piecewise Linear Model

x-score		parameter range of σ in piecewise linear model	WTP^d in piecewise linear model
$t + 0.5$	iff	$ts/(g + ts) \leq \sigma$	$ts/g \leq WPT^d$
$t - 0.5$	iff	$(t-1)s/[g + (t - 1)s] \leq \sigma < ts/(g + ts)$	$(t-1)s/g \leq WPT^d < ts/g$
...			
0.5	iff	$0 \leq \sigma < s/(g + s)$	$0 \leq WPT^d < s/g$
- 0.5	iff	$-s/(g - s) \leq \sigma < 0$	$-s/g \leq WPT^d < 0$
...	...	...	...
$-(t - 0.5)$	iff	$-ts/(g - ts) \leq \sigma < -(t-1)s/[g -(t-1)s]$	$-ts/g \leq WPT^d < -(t-1)s/g$
$-(t + 0.5)$	iff	$\sigma < -ts/(g - ts)$	$WPT^d < -ts/g$

t test-size parameter in the XY-test

s step-size parameter in the XY-test

g gap-size parameter in the XY-test

σ weight the decision maker puts on the passive person's payoff in the domain of disadvantageous inequality in the piecewise linear model

WTP^d for $WPT^d > 0$: $|WTP^d|$ = amount of own material payoff the decider is willing to give up in the domain of disadvantageous inequality in order to *increase* the other's material payoff by a unit
for $WPT^d < 0$: $|WTP^d|$ = amount of own material payoff the decider is willing to give up in the domain of disadvantageous inequality in order to *decrease* the other's material payoff by a unit (with inequalities reversed)

Note that the piecewise linear model implies piecewise constant willingness to pay (WTP) of the DM for income increases (or decreases) of the passive person (WTP = u_o/u_m, where the subscripts denote partial derivatives). In the domain of disadvantageous [advantageous] inequality we have $WTP^d = \sigma/(1 - \sigma)$ [$WTP^a = \gamma/(1 - \gamma)$, respectively]; if $\sigma \geq 0$ [$\gamma \geq 0$, respectively] then this term gives the own-money amount the DM is willing to give up in the domain of disadvantageous inequality [advantageous inequality, respectively] in order to *increase* the other person's material payoff by a single unit; symmetrically, if $\sigma < 0$ [$\gamma < 0$, respectively] then $-\sigma/(1- \sigma)$ [$-\gamma/(1 - \gamma)$, respectively] gives the own-money amount the DM is willing to give up in the domain of disadvantageous inequality [advantageous inequality, respectively] in order to *decrease* the other person's material payoff by a single unit. Thus, within the piecewise linear model x-scores translate into WTP^d as shown in the right-most column of Table 5 (again the translation for the y-score is similar except that WTP^d is replaced by WTP^a and strict inequalities are replaced by weak ones).

It is important to note that using estimates of the parameters of the piecewise linear model (or estimates of the piecewise constant WTP for changes in the income of the other) as a

cardinal metric for distributional preferences does not necessarily mean assuming piecewise linear preferences: In experimental set ups, where stakes tend to be small, the estimates are probably best interpreted as linear approximations of the true values. This interpretation is especially valid when the parameters of the piecewise linear model are estimated from the raw data using the McFadden (1974) random utility specification. In addition to yielding a cardinal metric that is comparable across studies, estimating the parameters of such a structural model has several other practical advantages as well:[32]

- As McFadden's random utility specification allows for noisy decisions, subjects with inconsistent choices do not have to be dropped. This may be crucial when a test design with high resolution (i.e. with large gap variable g and small step size s) is used, or when the binary choices are presented to the subjects one-at-a-time in random order (see Appendix A for a discussion of implementation issues).
- The parameter's standard errors enable statistical tests, for example, to check whether a subject's deviations from purely selfish behavior are statistically significant.
- The structural model could also be applied in the context of a finite mixture specification. This would allow to identify the prevalent social preference types and to endogenously classify each subject into the type that fits her behavior best.
- The parameters of the piecewise linear model can also be estimated when the test is applied in its multi-list variant where the (x, y)-score is no longer available.

6 Experimental Results Based on the Symmetric Basic Version of the Test

Here the data from a paper-and-pen experiment based on the symmetric basic version of the test is reported. The experiment was conducted in paper-and-pen (and several other design features reported below were applied) to convince subjects that neither other experimental subjects nor the experimenters could identify the person who has made any particular decision. This was done in an attempt to minimize the impact of experimenter demand and audience effects. See List (2007) for a discussion on experimenter demand effects and Hoffmann et al. (1994), Andreoni and Petrie (2004), and Andreoni and Bernheim (2009) for experimental evidence indicating that – depending on the experimental design – audience effects might have a large impact on subjects' behavior in dictator-game like situations.[33]

[32] I thank an anonymous colleague for suggesting this interesting discussion.

[33] See Hoffman et al. (1994) and Cox and Sadiraj (2012) for (almost double blind) experimental designs similar to the one employed here. Appendix A discusses alternative designs more suitable for the case where the *XY*-test is used as a tool in computer-supported experiments to address research questions in which distributional preferences play an important role, to control for subject pool effects, or to help to interpret data from other (unrelated) experiments. Those alternative protocols might be regarded as more problematic in terms of experimenter demand and audience effects but have advantages in other dimensions (an important one being that subjects are motivated

Experimental Procedures: Five experimental sessions were conducted manually (i.e., in pen-and-paper) at the University of Innsbruck in autumn 2009. Forty subjects who had not participated in similar experiments in the past were invited to each session using the ORSEE recruiting system (Greiner 2004). Since not all subjects showed up in time, 192 (instead of the invited 200) subjects from various academic backgrounds participated in total, and each subject participated in one session only. After arrival, subjects assembled in one of the two laboratories and individually drew cards with ID numbers (which remained unknown to other participants and the experimenters). Then instructions were distributed and read aloud.[34] Instructions informed subjects (i) that there are two roles in the experiment, the role of an 'active person' and the role of a 'passive person'; (ii) that there is exactly the same number of active and passive subjects in the experiment and that roles are assigned randomly; (iii) that each active person is matched with exactly one passive person and vice versa, and that at no point in time a participant will get to know anything regarding the identity of the person she/he is matched with; (iv) that active persons are called to make a series of ten binary decisions that determine not only their own earnings from the experiment but also the earnings of the passive person they are matched with; (v) that passive persons do not have a decision to make in the experiment and that their earnings will depend exclusively on the decisions of the active person they are matched with; (vi) that only one of the ten choice problems of each active person will be relevant for cash payments; and (vii) that cash payments could be collected the day after the experiment at one of the secretaries who handles also the cash payments for other experiments (to ensure that the amount a subject earns cannot be linked to her/his decisions). Then subjects were randomly assigned to one of the two roles; active persons stayed in the same room while passive persons were escorted to the adjacent laboratory.

In both rooms subjects were seated at widely separated computer terminals (computers were switched off) with sliding walls. Active persons were handed out a form consisting of two pages – an empty cover sheet and a decision sheet as described in the next paragraph – and they were asked to fill out the decision sheet in private. Passive persons received a form consisting of three pages – an empty cover sheet and a two-page questionnaire unrelated to the experiment – and they were asked to complete the questionnaire in private. After the tasks in both rooms had been completed, for each active person one of the choice problems was randomly selected via a manual device – a bingo ball cage handled by the active person – for the purpose of cash-payment generation. The payoff-relevant decision problem was written on the cover page of the active person and the person was given the opportunity to take in private a look at her/his

to think about their decisions carefully and make choices that reflect their true preferences even when the preference elicitation procedure is only one of several tasks in an experiment).

[34] *The instructions –not intended for publication– are in Appendix C.*

choice in the payoff-relevant decision problem. Now subjects in both rooms were asked to label (in private) the cover sheet of their document with their ID number. Then participants in both rooms were called to put their documents (again in private) in boxes before leaving the room. Anonymous cash payments started the next day – giving experimenters the opportunity to manually match active with passive persons in the meantime. Participants presented the card with their ID number to an admin staff person who did not know who did what for which purpose, nor how cash payments were generated and they got their earnings in exchange (the fact that cash payments would be made that way was clearly indicated in the instructions). On average subjects earned approximately 11 Euros plus a show up fee of 4 Euros.

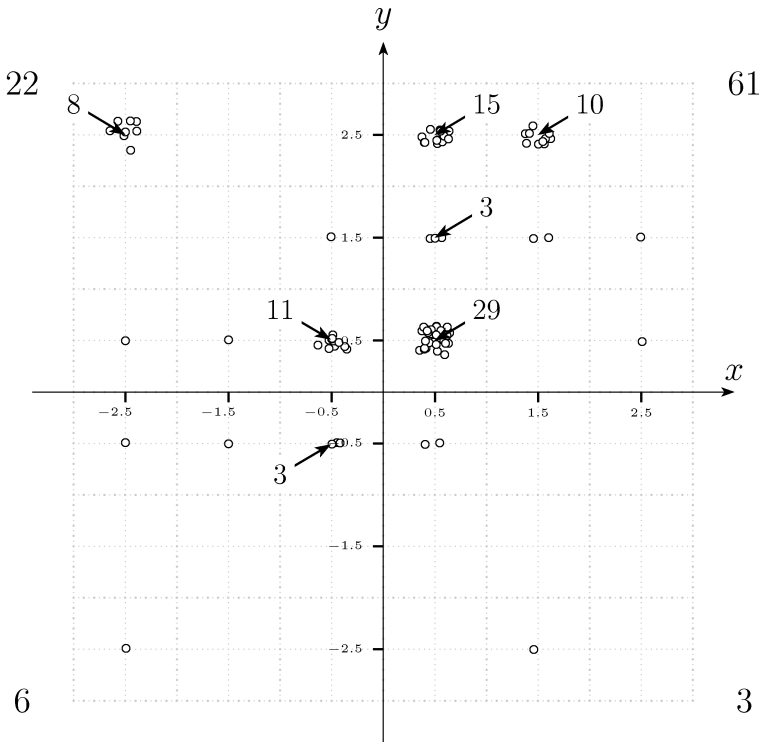
Experimental Design: The symmetric basic version of the test was implemented with $e = 10$, $g = 3$, $s = 1$, $t = 2$ and with experimental currency units corresponding to Euros. Thus, each active person (96 in total) was exposed to 10 binary decision problems with (10, 10) as the recurring equal-material-payoff allocation. The decision problems were presented in two tables, 5 in the *X*-Table (disadvantageous inequality) and 5 in the *Y*-Table (advantageous inequality). The design of the two tables was similar to that of tables 2 and 3.

Experimental Results: Of the 96 active subjects 4 (i.e., less than 5%) were eliminated in Step 1 of the procedure described in Subsection 3.3. The (x, y)-scores of the remaining 92 subjects were distributed as shown in Figure 7.[35] It is worth noting that more than half of the 36 points in the (x, y)-plane, where a subject could potentially sit, remain unoccupied, and only nine points are occupied by more than one subject. Thus, there is a sizeable amount of endogenous clustering. Also note that almost all subjects (87/92 = 95% of the population) reveal (weakly) more benevolent (less malevolent) preferences in the domain of advantageous than in the domain of disadvantageous inequality (i.e., their y-score exceeds the x-score). Taken together those two pieces of evidence (endogenous clustering of subjects and decisions consistent with convex preferences) indicate that subjects understand the binary choices presented to them and that the results reported here are driven by well-behaved distributional preferences and not by noise. The second piece of evidence also implies that non-convex types (most importantly, kick down and equality averse) are empirically irrelevant. Turning to convex types (convexity refers to the shape of indifference curves here), it is interesting to note that the behaviour of about two thirds of the subjects (those in the positive quadrant; 61/92 = 66.30% of the subject population) is consistent with altruistic preferences (there are only 2 subjects who reveal non-convex altruism), while the behaviour of (only) about one forth of the participants (those in the N/W quadrant; 22/92 = 23.92% of the subjects) is consistent with (any form of) inequality

[35] It is important to note that the data points in Figure 7 are jittered (to make each single point visible). For instance, the 29 observations scattered around the point (½, ½) all belong to the point (½, ½).

aversion.[36] Spiteful subjects (negative quadrant) exist, but they account for less than 7% of our population (and even spiteful subjects' score is consistent with convex preferences).

Figure 7: Absolute Frequency of (*x*, *y*)-Scores in Experiments Based on Basic Test Version
(96 active persons; 4 revealed inconsistencies; the figure is based on the remaining 92 subjects)



It is also interesting to observe that the behaviour of types at the border between altruism and inequality aversion ($x \in \{-½, ½\}$ and $y > 0$*;* 59/92 = 64.13% of the subject population) is consistent with maximin, while the behaviour of types at the border between inequality aversion and spite ($x < 0$ and $y \in \{-½, ½\}$; 18/92 = 19.57% of the population) is consistent with envy. Finally it is interesting to observe that the behavior of almost 50% of the population (those subjects with x and y in $\{-½, ½\}$; 45/92 = 48.91% of the population) is consistent with selfish preferences. Here note that the test assigns selfish subjects to one of the four quadrants in Figure 4 (Figure 7, respectively) according to their 'impartial distribution preferences' expressed in their choice behavior in the $(t+1)^{th}$ row of the two lists (where the DM decides between two allocations that differ only in the payoff of the passive person). For instance, a subject that is weakly benevolent in both domains gets $(x, y) = (½, ½)$, while a subject that is

[36] Given that differences in the distribution of types between subject pools are likely to be large this piece of evidence should not be interpreted as indicating that egalitarian motives are important only for a minority of subjects. See Fehr et al. (2006) for experimental evidence indicating that students (especially students of economics) are less egalitarian and more efficiency oriented than the rest of the population. See also the response by Engelmann and Strobel (2006) in the same issue.

weakly benevolent when ahead but weakly malevolent when behind gets $(x, y) = (-½, ½)$. Looking at Figure 7 we see that the choices of a majority of (but by far not the choices of all of) those subjects whose behavior is consistent with selfish preferences is also consistent with 'lexself' as defined by Fisman et al. (2007).[37]

7 Conclusions

This paper has proposed a geometric delineation of distributional preference types and a non-parametric approach for their identification in a two-person context. Major advantages of the proposed test over previous ones are (i) that it is *simple* and *short* as subjects' task is to make a small set of diagnostic choices without feedback; (ii) that it is *parsimonious* as it relies on a small set of primitive assumptions; (iii) that it is *general* as it directly tests the core features of different types of distributional preferences rather than concrete models or functional forms; (iv) that it is *flexible* as test size and test design can easily be fine-tuned to the research question of interest; (v) that it is *precise* as it identifies the archetypes of distributional concerns with arbitrary precision and also gives an index of preference intensity; and (vi) that it *minimizes experimenter demand effects* as subjects are asked to make binary decisions in a neutral frame and do not have the option to do nothing.[38] Those features together suggest that the proposed test might be suitable as a tool in experimental economics to disentangle the impact of distributional preferences from that of other factors thereby helping to interpret data from other (unrelated) experiments (similar to the choice list tests used to elicit risk attitudes; see Holt and Laury 2002, or Dohmen et al. 2010 and 2011).[39]

That the proposed test is indeed suitable for that purpose has been shown in two recent studies: Balafoutas et al. (2012) investigate in a standard lab experiment the relationship between distributional preferences and competitive behavior and find (a) that distributional archetypes (as assigned by the proposed test) differ systematically –and in an intuitively plausible way– in their response to competitive pressure, in their performance in a competitive environment and in their willingness to compete; and (b) that controlling for the effects of distributional preferences, as well as for risk attitudes and some other factors, closes the large gender gap in competitive behavior found in earlier studies (by Niederle and Vesterlund 2007 and 2010, for instance). This is an important finding because it indicates that the gender gap in competitiveness is largely driven by mediating factors (potentially accessible to policy

[37] Here note that in the *XY*-test a subject with lexself preferences necessarily ends up with an (x, y)-score of $(½, ½)$, independently of the parameterization of the test. As can be seen in Figure 7 about one third of the participants (specifically, 29 of the 92 classified subjects) ended up with such a score.

[38] This is in contrast to the standard dictator game which gives the DM a windfall gain and then the option to share. This makes it pretty clear what would be considered decent behavior by the experimenter.

[39] Some implementation issues relevant for this purpose are discussed in Appendix A.

intervention) and not by gender per se. Hedegaard et al. (2011) examine in a large-scale internet experiment the impact of distributional concerns on the contribution behavior in a standard (linear) public goods game and find (a) that distributional archetypes differ systematically –and in an intuitively plausible way– in their contribution behavior; and (b) that accounting for the differences explains roughly half of the gap between actual behaviour of subjects in the lab and the theoretical benchmark derived under the assumption that players are rational and selfish (and that this fact is common knowledge). Again, this is an important finding because it helps to disentangle the impact of distributional concerns on the behavior of subjects in social dilemma games from that of other factors – as beliefs on others' behavior or intentions, for instance. Together the findings in those studies clearly indicate that associating subjects with one of the proposed archetypes of distributional concerns has explanatory value and that the proposed test is indeed a valid control instrument in experimental economics.

Beyond its potential to act as a control instrument in experimental economics, other potentially fruitful applications of the test include *(a) investigating the stability of distributional preferences over different domains* (for instance, a potential shortcoming of the approach proposed here is its focus on the two-agents case; investigating whether the preferences revealed in that context carry over to a richer environment is surely an important issue);[40] (b) *investigating possible links between distributional preferences and other forms of other-regarding preferences* (for instance, "Are altruists more or less likely to be motivated by positive or negative reciprocity?", "Do altruism and altruistic rewarding (or altruistic punishment) go together or are they mutually excluding ways to reach the same goal – promoting private provision of public goods?" [41], or "Is the test-based classification of subjects in distributional-preference types somehow correlated with the propensity to be motivated by trust?");[42] and (c) *applying the proposed test* (together with tests for risk and time preferences and for personality traits) *in experiments with large demographic variation* (age, gender, income, education) *or with a representative sample of the population to detect patterns and correlations* (for instance, "Are distributional preferences and risk attitudes or time preferences

[40] In this regard the *XY*-test is very similar to the standard risk-attitude elicitation procedures that ask subjects to compare a binary lottery either to another binary lottery (as in Holt and Laury 2002) or to a risk-free alternative (as in Dohmen et al. 2010 and 2011). In both domains, the risk-preference domain and the distributional-preference domain, the hope is that the preference type revealed in the binary environment is informative of the attitude of the decision maker in richer environments.

[41] Altruistic rewarding (punishment) is the propensity to give rewards to (impose sanctions on) others for 'norm-abiding-behavior' ('norm-violating-behavior') even if rewards (sanctions) are costly for the rewarder (punisher) and yield no private material benefits whatsoever. See Fehr and Gächter (2000b and 2002), or Fehr and Fischbacher (2004), for studies investigating the power of altruistic rewards and punishments.

[42] At first blush this avenue of research seems to be closely related to that of Blanco et al. (2011). It is not that close, however, since the research question proposed here is not whether the behavior of players in strategic games can be explained by distributional preferences (alone) but rather whether there is a link between distributional preferences and other forms of other-regarding preferences.

somehow related?", "Are there gender differences in the distribution of archetypes?"[43] or "What is the impact of age and income on distributional preferences?"). Beyond economics the proposed test might help to address important research questions in biology and psychology as, for instance, "What determines human altruism (or spite)?" or, "What drives altruistic punishments and rewards?".

For those and many other interesting research questions identification of distributional preference types in a "clean" environment appears to be a natural first step. The proposed non-parametric approach seems to be well suited for this purpose. Turning back to the quote at the start of the paper the hope is that it turns out to be "as simple as possible, but not one bit simpler".

[43] This research question is not new, of course – see Andreoni and Vesterlund (2001), for instance. The proposed approach allows for a more detailed investigation of this issue, though. And it is suitable to be integrated in large-scale experiments (as shown by Hedegaard et al. 2011).

Acknowledgements: I am grateful to Uwe Dulleck, Ernst Fehr, Anita Gantner, Morten Hedegaard, Steffen Huck, Ricci Hule, Georg Kirchsteiger, Rupert Sausgruber, Matthias Sutter and Jean-Robert Tyran for helpful suggestions and comments, and to Dominik Erharter and Tanja Hörtnagel for excellent research assistance. I also thank the audience at the 2010 Thurgau Experimental Economics Meeting as well as seminar participants at the University of Zürich for helpful comments. Part of this research was conducted while I was visiting the Queensland University of Technology and I thank its School of Economics and Finance for its great hospitality. Financial support from the Austrian Science Foundation (FWF) through grant numbers P-20796 and P-22669 and from the Austrian National Bank (OeNB Jubiläumsfonds) through grant number 13602 is gratefully acknowledged.

References

Andreoni, J. and Bernheim, D. (2009). Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica* 77, 1607–1636.

Andreoni, J. and Miller, J. (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica* 70, 737-753.

Andreoni, J. and Petrie, R. (2004). Public Goods Experiments Without Confidentiality: A Glimpse into Fund-Raising. *Journal of Public Economics* 88, 1605-1623.

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

Balafoutas, L., Kerschbamer, R. and Sutter, M. (2012). Distributional Preferences and Competitive Behavior. *Journal of Economic Behavior and Organization* 83, 125-135

Battigalli, P. and Dufwenberg, M. (2007). Guilt in Games. *American Economic Review, Papers and Proceedings* 97, 170-176.

Becker, G. (1974). A Theory of Social Interactions. *Journal of Political Economy* 82, 1063-93.

Binmore, K. (2005). Natural Justice, *Oxford University Press*, Oxford, UK

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

Bolle, F. and Kritikos, A. (2006). Reciprocity, Altruism, Solidarity: A Dynamic Model. *Theory and Decision* 60, 371-394.

Bolton, G. (1991). A Comparative Model of Bargaining: Theory and Evidence. *American Economic Review* 81, 1096-1136.

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

Brandts, J., Riedl, A. and van Winden, F. (2009). Competitive Rivalry, Social Disposition, and Subjective Well-Being: An Experiment. *Journal of Public Economics* 93, 1158-1167.

Brosig, J. (2002). Identifying Cooperative Behavior: Some Experimental Results in a Prisoner's Dilemma Game. *Journal of Economic Behavior and Organization* 47, 275–290.

Bruhin, A., Fehr-Duda, H. and Epper, Th. (2010). Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Econometrica* 78, 1375-1412.

Cabrales, A., Miniaci, R., Piovesan, M. and Ponti, G. (2010). Social Preferences and Strategic Uncertainty: An Experiment on Markets and Contracts. *American Economic Review* 100, 2261-2278.

Camerer, C. (1997). Progress in Behavioral Game Theory. *Journal of Economic Perspectives* 11, 167–188.

Camerer, C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. *Princeton University Press*, Princeton, USA.

Charness, G. and Dufwenberg, M. (2006). Promises and Partnership. *Econometrica* 74, 1579-1601.

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

Cox, J., Friedman, D. and Gjerstad, S. (2007). A Tractable Model of Reciprocity and Fairness. *Games and Economic Behavior* 59, 17-45.

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

Cox, J. and Sadiraj, V. (2012). Direct Tests of Individual Preferences for Efficiency and Equity. *Economic Inquiry* 50, 920-931.

Cox, J., Sadiraj, K. and Sadiraj, V. (2008). Implications of Trust, Fear, and Reciprocity for Modeling Economic Behavior. *Experimental Economics* 11, 1-24.

Dawes, Ch., Fowler, J., Johnson, T., McElreath, R. and Smirnov, O. (2007). Egalitarian Motives in Humans, *Nature* 446, 794-796.

Dohmen, Th., Falk A., Huffman, D. and Sunde, U. (2010) Are Risk Aversion and Impatience Related to Cognitive Ability? *American Economic Review* 100, 1238-1260.

Dohmen, Th., Falk A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G. (2011) Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences. *The Journal of the European Economic Association* 9, 522-550.

Dufwenberg, M. and Gneezy, G. (2000). Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior* 30, 163-182.

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

Duesenberry, J. (1949). Income, Saving and the Theory of Consumer Behavior. *Harvard University Press*, Cambridge, MA.

Engelmann, D. and Strobel, M. (2004). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *American Economic Review* 94, 857-869.

Engelmann, D. and Strobel, M. (2006). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Reply. *American Economic Review* 96, 1918-1923.

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

Fehr, E., Bernhard, H., and Rockenbach, B. (2008). Egalitarism in Young Children. *Nature* 454, 1079-1083.

Fehr, E., and Fischbacher, U. (2003). The Nature of Human Altruism. *Nature* 425, 785-791.

Fehr, E. and Fischbacher, U. (2004). Third-Party Punishment and Social Norms. *Evolution and Human Behavior* 25, 63-87.

Fehr, E. and Gächter, S. (2000a). Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives* 14, 159-181.

Fehr, E. and Gächter, S. (2000b). Cooperation and Punishment in Public Goods Experiments. *American Economic Review* 90, 980-994.

Fehr, E. and Gächter, S. (2002). Altruistic Punishment in Humans. *Nature* 415, 137-140.

Fehr, E., Kirchsteiger G. and Riedl, A. (1998). Gift Exchange and Reciprocity in Competitive Experimental Markets. *European Economic Review* 42, 1-34.

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

Fehr, E. and Schmidt, K. (2006). The Economics of Fairness, Reciprocity and Altruism: Experimental Evidence. In: S. Kolm and J.-M. Ythier (eds.) *Handbook of The Economics of Giving, Altruism and Reciprocity*, Elsevier, 1.ed., vol. 1.

Fehr, E., Naef, M. and Schmidt, K. (2006). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment. *American Economic Review* 96, 1912-1917.

Fershtman, Ch., Gneezy, U. and List, J. (2012). Equity Aversion: Social Norms and the Desire to be Ahead. *American Economic Journal Micro,* forthcoming

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

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

Greiner, B. (2004). The Online Recruiting System ORSEE 2.0 - A Guide for the Organisation of Experiments in Economics. *WP Series in Economics 10,* University of Cologne.

Griesinger, D. and Livingston, J. (1973). Toward a Model of Interpersonal Motivation in Experimental Games. *Behavioral Science* 18, 173-188.

Hedegaard, M., Kerschbamer, R. and Tyran, J.-R. (2011). Correlates and Consequences of Distributional Preferences: An Internet Experiment. *Mimeo,* Department of Economics, University of Copenhagen.

Hennig-Schmidt, H. (2002). The Impact of Fairness on Decision Making - An Analysis of Different Video Experiments. In: F. Andersson and H. Holm (eds.), Experimental Economics: Financial Markets, Auctions, and Decision Making, *Kluwer Academic Publishers*, Boston, 185-210.

Hoffman, E., McCabe, K., Shachat, K. and Smoth, V. (1994), Preferences, Property Rights, and Anonymity in Bargaining Games, *Games and Economic Behavior* 7, 346-380.

Holt, C. and Laury, S. (2002). Risk Aversion and Incentive Effects. *American Economic Review* 92, 1644-1655.

Iriberri, N. and Rey-Biel, P. (2013). Elicited Beliefs and Social Information in Modified Dictator Games: What do Dictator Believe Other Dictators Do? *Quantitative Economics,* forthcoming.

Kirchsteiger, G. (1994). The Role of Envy in Ultimatum Games. *Journal of Economic Behavior and Organization* 25, 373-89.

Konow, J. (1993). Which Is the Fairest One of All? A Positive Analysis of Justice Theories. *Journal of Economic Literature* 41, 1188-1239.

Kuziemko, I., Buell, R., Reich, T. and Norton, M. (2011). "Last-Place Aversion": Evidence and Redistributive Implications. *NBER Working Paper* 17234, National Bureau of Economic Research, Cambridge, Massachusetts.

Levine, D. (1998). Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics* 1, 593-622.

Liebrand, W. (1984). The Effect of Social Motives, Communication and Group Sizes on Behavior in an N-Person Multi Stage Mixed Motive Game. *European Journal of Social* Psychology 14, 239-264.

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

Malmendier, U. and Szeidl, A. (2008). Fishing for Fools. *Mimeo*, UC Berkeley

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In Zarembka, P. (ed.) (1974*). Frontiers in Econometrics*. Academic Press, New York.

Mui, V. (1995). The Economics of Envy. *Journal of Economic Behavior and Organization* 26, 311-336.

Niederle, M. and Vesterlund, L. (2007) Do Women Shy Away From Competition? Do Men Compete Too Much? *Quarterly Journal of Economics* 122, 1067–1101.

Niederle, M. and Vesterlund, L. (2010) Explaining the Gender Gap in Math Test Scores: The Role of Competition. *Journal of Economic Perspectives* 24, 129–144.

Offerman, T., Sonnemans, J. and Schram, A. (1996). Value Orientations, Expectations and Voluntary Contributions in Public Goods, Economic Journal 106, 817-845.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review* 83, 1281-1281.

Rawls, J. (1971). A Theory of Justice. Belknap Press of Harvard U. Press, Cambridge, USA

Sobel, J. (2005). Interdependent Preferences and Reciprocity. *Journal of Economic Literature* 43, 392-436.

Sonnemans, J., Schram, A. and Offerman, T. (1998). Public Good Provision and Public Bad Prevention: The Effect of Framing. *Journal of Economic Behavior and Organization* 34, 143-161.

Sutter, M., Haigner, S. and Kocher, M. (2010). Choosing the Stick or the Carrot? - Endogenous Institutional Choice in Social Dilemma Situations. *Review of Economic Studies* 77, 1540-1566.

Tyran, J.-R. and R. Sausgruber (2006). A Little Fairness May Induce a Lot of Redistribution in Democracy. *European Economic Review* 50, 469-485.

Van Dijk, F., Sonnemans, J. and van Winden, F. (2002). Social Ties in a Public Good Experiment. *Journal of Public Economics* 85, 275-299.

Appendix (For Online Publication)

This appendix consists of three parts. Parts A and B discuss two themes that seem important when the test is used as a tool in experimental economics to answer specific research questions in which distributional preferences play an important role, to control for subject pool effects, or to help to interpret data from other (unrelated) experiments: While Appendix A is devoted to a discussion of implementation issues for this case, Appendix B contains a more detailed discussion of the ring-test, probably the most important contender in this domain. Appendix C contains the instructions to the experiments reported in Section 6 of the paper.

Appendix A: Implementation Issues

While the non-parametric identification approach proposed in the body of the paper seems in principle well suited as a tool in experimental economics to be added to arbitrary (other) experiments, there are several practical issues that need to be addressed.

Role Assignment: At least 3 different protocols regarding role assignment have been used in the literature on elicitation of distributional preferences, *fixed role assignment***,** where roles (active DM and passive person) are assigned ex ante, and only active DMs decide while passive persons do nothing (see, e.g., Cox, Sadiraj and Sadiraj 2008, and Cox and Sadiraj 2012); *role uncertainty***,** where each subject decides in the role of the active DM, and only later subjects get to know whether their decision is relevant – i.e., whether they have been chosen as DM or as passive person (this procedure was used by Engelmann and Strobel 2004 and by Blanco, Engelmann, and Normann 2011); and *double role assignment*, where each subject decides, and each subject gets two payoffs, one as an active DM and one as a passive person (as in Andreoni and Miller 2002, Anderoni and Vesterlund 2001 and in Fisman, Kariv, and Markovits 2007).[44] While the fixed rule assignment (this protocol has been used to produce the data presented in body of the paper) seems to be the cleanest procedure from a theoretical point of view, it is not practicable when the test is intended as a tool to be added to arbitrary other experiments (since it would imply inviting twice as many subjects than needed for the main treatments). Each of the other two protocols seems to have some drawbacks. Consider the role uncertainty protocol first. Since it introduces an element of randomness in the determination of the payoff allocation

[44] Actually, the double role assignment protocol comes in two varieties; while in version 1 (the version discussed in the main text) the computer program makes sure that a subject's active DM is a different participant than a subject's passive person (and instructions are very explicit about this), version 2 (often used in the implementation of the ring-test and in related tests designed by social psychologists) has fixed pairs, meaning that a subject's active DM is the same participant as her passive person. Version 2 seems theoretically problematic and is therefore ignored in the discussion in the main text.

resulting from a decision it raises theoretical questions related to the issue process fairness vs. outcome fairness (see Andreoni and Bernheim 2009 for evidence that some subjects care for process fairness). Secondly (and related to the first point), expectations about the behavior of the passive person in the counterfactual situation where she is the active DM might influence choices (provided process fairness matters; if not then not).[45] The double rule assignment protocol seems to be better in the former dimension, but it might be worse in the latter as a subject's expectations about what she gets as a passive person and about what her passive person gets as the active DM are even more likely to enter the picture. In sum, both double rule assignment and role uncertainty protocol have their own problems and it is ultimately an empirical question which one performs better (in predicting the decisions in other distributional tasks, for instance). Here, promising evidence in support of the role uncertainty protocol is provided by Hedegaard et al. (2011). In their large-scale internet experiment they employ both the fixed role assignment and the role uncertainty protocol in a between subjects design and they show that the two protocols yield results that are statistically indistinguishable, both regarding the distribution of archetypes they yield and regarding the ability to predict behavior in other games.[46]

Presentation of the Binary Decision Problems: In the paper-and-pen experiments reported in the body of the paper the binary decision tasks were presented to the subjects in *ordered lists* or tables (similar to the lists often used in risk-attitude elicitation tasks). In computer-aided experiments (using z-Tree developed by Fischbacher 2007, for instance) presenting the binary decisions *one-at-a-time in random order* (i.e., each binary decision on an own screen) might be an attractive alternative. Experience with both presentation techniques in computer-based experiments (where the test was added as a control at the end of the main experiments) suggests that the randomized test version produces (slightly) more inconsistencies (more than one switch in at least one of the two lists; or switches in the wrong direction) but might (slightly) increase the predictive power of the test for the classified subjects. This indicates that especially when the test is added as a control at the end of other experiments the presentation of the binary

[45] If subjects have distributional preferences in the textbook variety (only outcomes matter) then expectations should not shape decisions in the role uncertainty protocol: with some probability α (= ½) the other person is the active DM und your expectations about her/his behavior influence what you expect to get in that case; given your preferences and expectations this yields a fixed "utility" that you get with probability α; to maximize your overall (expected) utility you still have an incentive to maximize that part of your "utility" that realizes with probability 1 - α. No expectations enter in that part of your overall expected utility. The story changes if (some) subjects are not (only) concerned with the "fairness" of outcomes but (also) with the "fairness" of lotteries. This is beyond pure distributional preferences, though.

[46] This is in line with earlier evidence provided by Engelmann and Strobel (2004) who find relatively small and insignificant differences in the choices of subjects between their main treatment with role uncertainty and their control with fixed role assignment.

decision tasks might be critical in recovering reliable data on distributional preferences. The simplest procedure (presenting the binary decision tasks in a table) might not produce the most reliable results then, the main reason being that subjects do not think carefully enough about the tasks when they are presented all at once in a table.[47] A "third way" to implement the test in the lab is to present the binary choices first in a totally randomized way (i.e., also randomized across blocks and in the presentation of the recurring alternative on the left or the right hand side), and to show subjects the ordered lists (with their decisions in the different choice tasks) when they are done with all choices. They can then revise their choices if they like.[48] The latter procedure forces players to rethink their decisions; this might help to get fewer inconsistencies than in the random order design and might at the same time yield higher predictive power than the list versions. Again, it is ultimately an empirical question to sort out what the best way is to present the choice problems to the subjects when the test is used as a tool in experimental economics.

Clustering of Subjects: Clustering means dividing subjects into groups (clusters) so that members of one group are somehow similar to each other and dissimilar from members of other groups. There are many ways how clustering might be performed after having assigned to each subject an (x, y)-score. An obvious one is to cluster subjects into 9 groups corresponding to the 9 archetypes described in the body of the paper. This will often not meet the needs of the experimenter, though. The preferred way to group subjects will rather depend on at least three factors: (i) on the number of subjects taking part in the experiment; (ii) on the test version used; and (iii) on the research question under investigation.

Number of Subjects: If the number of subjects taking part in a study is small it does not make much sense to group them into too many clusters. Several approaches to divide subjects into 2-4 clusters spring to mind and the preferred one will, in general, depend on the other two dimensions discussed below. For instance, when a rough test version (high value of the quotient *s/q;* see the discussion in the next paragraph) is used a natural approach to divide subjects into 3-4 clusters is to use the sign of the two scores as the discriminator (both scores positive: altruistic; x-score negative, y-score positive: inequality averse; both scores negative: spiteful; x-score positive, y-score negative: equality averse; the latter class will be almost empty, though), while a finer test version would suggest that subjects with x- and y-score in {-0.5, +0.5} should be grouped in an own cluster (egoistic). On the other extreme, if the number of subjects taking

[47] Misclassification of subjects who do not think carefully enough about the alternatives in a binary decision task seems less of an issue in the experiments reported in the body of the paper because subjects' task was merely to make 10 binary decisions there. We therefore opted for a design that is cleaner in terms of experimenter demand and audience effects.

[48] Hedegaard et al. (2011) employ this protocol in their internet experiments.

part in the study is large there might be no need for exogenous clustering at all. For instance, in the experiments reported in Section 6 of the paper of the 36 points in the (x, y)-pane where subjects can potentially sit, more than half remained unoccupied, and only on nine points there was more than one subject sitting. Here it might make sense to work with nine clusters or less and to assign subjects who sit alone (or almost alone) at a point to one of the more frequented adjacent points according to some distance measure.

Test Version: One important design decision in the symmetric basic version of the test is the choice of the quotient s/g since this quotient determines the precision with which egoistic subjects are identified. If egoistic subjects are identified with high precision it might make sense to work with the following 5 clusters (where the last one is empty with high probability): $x \in \{-0.5, +0.5\}$ and $y \in \{-0.5, +0.5\}$: egoistic; $x \geq 0.5$ and $y \geq 0.5$ and at least one inequality strict: altruistic (or efficiency loving); $x \leq 0.5$ and $y \geq 0.5$ and at least one inequality strict: inequality averse; $x \leq 0.5$ and $y \leq 0.5$ and at least one inequality strict: spiteful; and $x \geq 0.5$ and $y \leq 0.5$ and at least one inequality strict: equality averse.

Research Question: The research question under investigation is important for clustering, of course. For instance, for predicting the behavior of a subject in a standard dictator game, only the y-score should be important. So, for predicting behavior in dictator game like situations it might be sensible to divide subjects only according to this dimension (for instance, into 3 clusters, one with $y<0$, the second with $y=0.5$ and the third with $y>0.5$; or, depending on the test design, into 4 or more clusters by splitting up the $y>0.5$ group in subgroups).

Appendix B: The Ring-Test

The ring-test has been developed by social psychologists (standard references are Griesinger and Livingston 1973 and Liebrand 1984) to assess "social value orientations" (SVO) at the individual level. In the standard implementation of this test subjects are asked to make $4k$ binary choices between (m, o) allocations taken from a fixed set with $4k$ elements, where k is some positive integer and where m stands for the DM's and o for the passive person's material payoff (as before). Typical values for k are 6 or 7 implying 24 or 32 binary choices. The $4k$ alternatives in the set lie equally spaced on a circle with the origin of the (m, o) plane serving as the centre. The radius of the circle is r, say. The facts that (i) there are $4k$ alternatives, that (ii) alternatives are located equally spaced on the circle and that (iii) k is an integer together imply that if (m, o) is one of the alternatives used in the test, then $(m, -o)$, $(-m, o)$ and $(-m, -o)$ are other alternatives used in the test. This property is important as it is used heavily in the typical identification procedure. In each of the $4k$ binary choice tasks the individual is asked to decide

between two adjacent (m, o) alternatives on the circle.[49] Since any (m, o) alternative has two immediate neighbours, it is presented in two different choice tasks to each subject.

Social psychologists have developed a standard procedure to identify distributional type and intensity of distributional concerns on the basis of observed choices in this test. The procedure is based on the assumption that preferences of subjects can be represented by a utility or motivational function of the form

$$u(m, o) = \mu m + \lambda o, \qquad (rt)$$

where μ and λ are two a priori unrestricted parameters (that is, no assumption guaranteeing that $\mu + \lambda = 1$ – or that at least one of the parameters is strictly positive, or whatsoever – is made; more on this below), which are assumed to be constant. Utility or motivational functions of this form imply linear and parallel indifference curves in the (m, o) space.[50] For instance, if both parameters are strictly positive, we get negatively sloped indifference curves, similar to those we have attributed to our altruists. Similarly, if λ is negative while μ is positive, we get positively sloped indifference curves, similar to the ones we have attributed to competitive subjects. Note that our strict m-monotonicity assumption applied to this function simply amounts to $\mu > 0$. No such assumption is typically imposed by social psychologists.[51]

Given a motivational or utility function of the form (rt) a subject has a unique most preferred point on each circle in the (m, o) space. Denote this point by (m^*, o^*). For $\mu > 0$ the point (m^*, o^*) is found by searching for the indifference curve that is tangential to the right hand side of the circle (because for $\mu > 0$, points to the right of a given indifference set are strictly preferred to points in the set), for $\mu < 0$ the point (m^*, o^*) is found by searching for the indifference curve that is tangential to the left hand side of the circle (because for $\mu < 0$ points to the left of a given indifference set are strictly preferred to points in the indifference set), for $\mu = 0$ (implying horizontal indifference curves) the point (m^*, o^*) lies either at (0, r) or at (0, -

[49] Since on the circle we have $m^2 + o^2 = r^2$, the sum $m + o$ is not constant. The sum is maximal, where m and o are positive and $m=o$ and minimal where m and o are negative and $m=o$.

[50] Along an indifference "curve" we have $u(m, o) = c$ for some constant c. Thus, for $\lambda \neq 0$ the indifference curve associated with "utility level" c is given by o = $(c-\mu m)/\lambda$, while for $\lambda = 0$ the indifference curve is a vertical line at $m=c/\mu$. Other indifference curves are simply parallel displacements of the indifference curve associated with c. In other words, linear preferences as described by equation (rt) have the property that the consumer's entire preference relation can be deducted from a single indifference set (or "indifference curve").

[51] For instance, social psychologists allow μ to be negative and λ to be positive, yielding indifference curves that have positive slope (similar to the ones the body of the paper has attributed to spiteful agents) but where the upper contour sets are to the left of the indifference curves (if μ is not too negative an individual with such preferences is classified by the ring-test as a "SVO altruist"; see the discussion below and Figure A1 in this appendix for typical indifference curves). They also allow both parameters to be negative, resulting in negatively sloped indifference curves (similar to the ones the body of the paper has attributed to altruists) but have the upper contour set again to the left (in a certain range, such individuals are classified by the ring-test as "SVO aggressors"; again, see Figure A1 in this appendix).

r) depending on whether $\lambda > 0$ (in this case points above a given indifference set -which is a horizontal line, as said- are strictly preferred to points in the indifference set) or $\lambda < 0$ (in this case points below a given indifference set -which is again a horizontal line- are strictly preferred to points in the indifference set). Furthermore, in the choice between two adjacent points on the circle subjects deciding such as to maximize a motivational or utility function of the form (*rt*) always select the alternative closest in distance to the point (*m**, *o**). This is so, because the alternative which is closer in distance to (*m**, *o**) is necessarily on a higher indifference curve (remember that indifference curves are parallel displacements of each other).

Consider a perfectly rational subject with a motivational or utility function of the form (*rt*). Suppose her most preferred point on the circle, (*m**, *o**), is among the alternatives compared in the ring-test. [If it is not, among the available alternatives take the one closest in distance to (*m**, *o**) -and slightly abusing notation- denote it (*m**, *o**).] Then in the 4*k* pair-wise comparisons the subject will select all alternatives but (*m**, *o**) and (-*m**, -*o**) exactly once; this is so because in one and only one of the two comparisons with a neighbour each of these alternatives has the shortest distance to (*m**, *o**). Furthermore, the point (*m**, *o**) is selected exactly twice (because it is strictly better than both of its neighbours) and the point opposite to (*m**, *o**), i.e., the point (-*m**, -*o**), is never selected (because it is worse than any other point from the subject's perspective). Moreover, by adding up the 4*k* selected payoff vectors separately for *m* and *o*, each of the 4*k*-*2* alternatives selected exactly once cancels out with the alternative on the opposite side of the circle, and we are left with the vector (2*m**,*2o**), which is consistent with exactly one motivational function of the form (*rt*).[52] Here note that the above described procedure generically is only clean for linear indifference curves. *Any* other shape of indifference curves might potentially yield problems.

Of course, real subjects in experiments do not always make choices that are exactly consistent with maximizing a function of the form (*rt*). Then the summation vector (resulting from summing up all the choices of a subject) divided by 2 is not a point on the circle. Social psychologists then lay a ray starting from the origin through the summation vector and take the *(m, o)* point at the intersection between the ray and the circle as the most preferred point of the subject on the circle. The distance between that point and half the summation vector is then taken as a measure of inconsistency of choices (with a motivational function of the form (*rt*)).

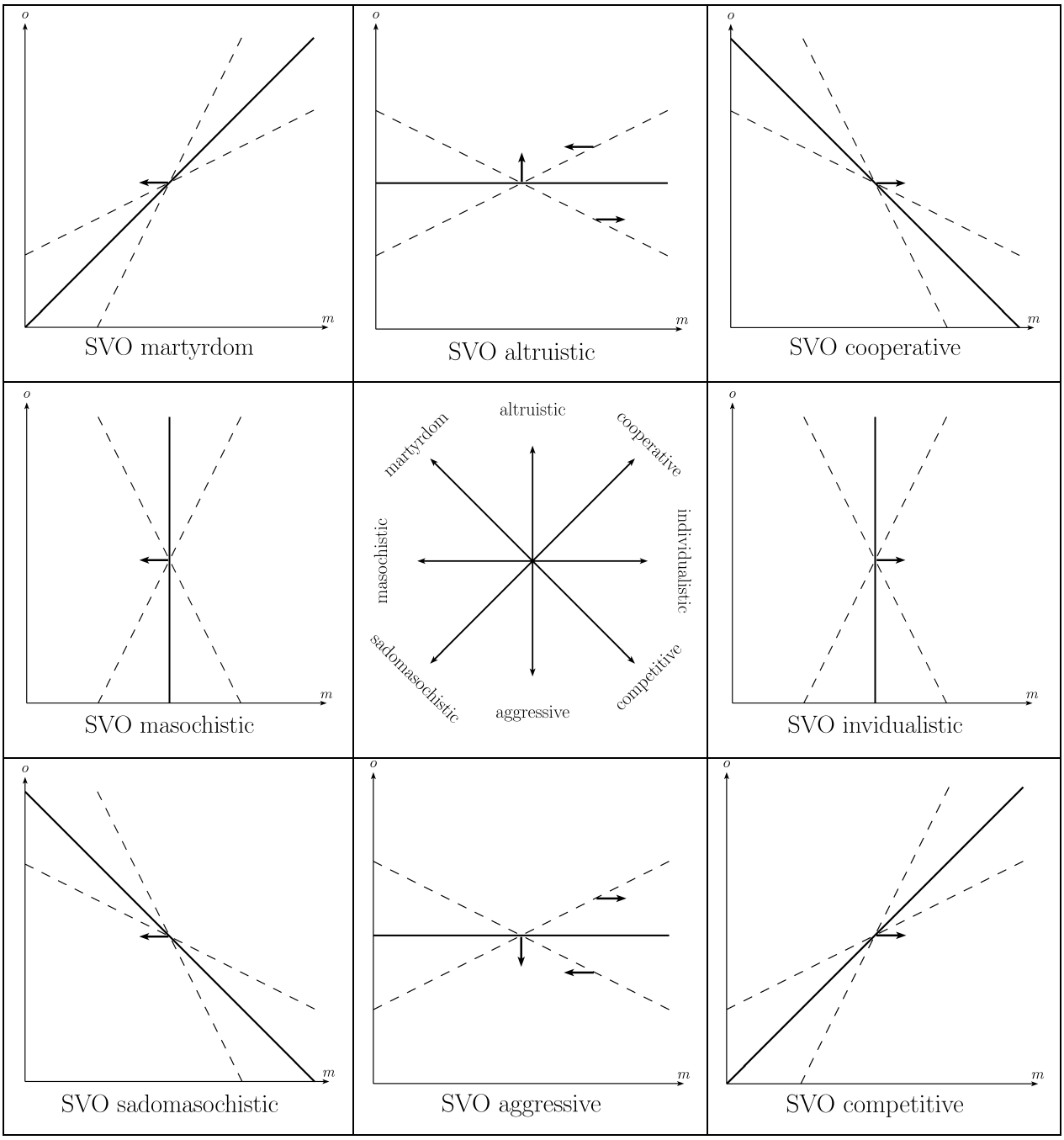
Subjects are then either classified into the 5 "*social value orientation"* (*SVO*) *types* altruistic, cooperative, individualistic, competitive and aggressive, or into 8 clusters, the 5 listed plus sadomasochistic, masochistic and martyrdom (see Figure A1 at the end of this appendix).

[52] As is easily verified, the described "technique" for finding the point (*m**, *o**) from an indifference curve map can easily be reversed: simply take the tangent to the circle at the point (*m**, *o**) to get the relevant indifference curve; then solve for the associated utility function along the lines described in Footnote 50 above.

The prototypical "*SVO altruist*" has horizontal indifference curves with upper contour sets in direction north; other indifference curve slopes consistent with SVO altruism are slightly negative slope with upper contour sets to the right and slightly positive slope with upper contour sets to the left. In terms of rays from the origin, rays between degree 67.5° and 112.5° are classified as SVO altruists.[53] The procedure in the body of the paper would classify SVO altruists with most preferred point between 67.5° and 90° as altruistists (they have negatively sloped indifference curves and upper contour sets to the right), while SVO altruists with most preferred point between 90° and 112.5° violate strict *m*-monotonicity and are therefore outside the admitted range of preferences (those types are perhaps more interesting for psychologists than for economists – in terms of the motivational function (*rt*) such subjects decide as if $\mu<0$ and $\lambda>0$). The prototypical *"SVO cooperator"* has negatively sloped indifference curves, the *XY*-test would classify such subjects as altruists, too. In terms of rays from the origin, rays between degree 22.5° and 67.5° are classified as SVO cooperators. The prototypical "*SVO egoist"* (named "individualistic" in Figure A1) has vertically sloped indifference curves with upper contour sets to the right; this classification is completely in line with the *XY*-test classification. To have classes of equal size (in terms of distance on the circle, not necessarily in terms of number of subjects in a class), SVO egoists are not restricted to a single point (ray with 0°) but extend from -22.5° (implying indifference curves with a slightly positive slope) to +22.5° (implying negatively sloped indifference curves). The prototypical *"SVO competitor"* has positively sloped indifference curves, the *XY*-test would classify such subjects as competitors, too. In terms of rays from the origin, rays between degree -67.5° and -22.5° are classified as SVO competitors. The prototypical "*SVO aggressor*" has horizontal indifference curves with upper contour sets in direction south; other indifference curve shapes consistent with SVO aggressors are slightly negative slope with upper contour sets to the left and slightly positive slope with upper contour sets to the right. In terms of rays from the origin, rays between degree -112.5° and -67.5° are classified as SVO aggressive. The *XY*-test would classify SVO aggressors with most preferred point between -67.5 and -90 as competitors (they have positively sloped indifference curves and upper contour sets to the right), SVO aggressors with most preferred point between -90° and -112.5° fall outside the considered range of preferences and are perhaps again more interesting for psychologists than for economists (in terms of the motivational function (*rt*) such subjects decide as if $\mu<0$ and $\lambda<0$). See Figure A1 below for a characterization of the types "*SVO sadomasochist*", "*SVO masochist*" and "*SVO martyr*".

[53] When discussing angles of rays the reference ("the 0°") is always the horizontal line corresponding to vertical indifference curves with upper contour sets to the right. Here note that indifference curves associated with a given ray are orthogonal to the ray.

Figure A1: Indifference Curves of Ring-Test Social Value Orientation (SVO) Types



arrows indicate the locus of the upper contour set

Appendix C: Instructions (Translated from German)

Welcome and thank you for participating!

You are taking part in an economic experiment on decision making. A research foundation has provided the funds for conducting the experiment. You can earn a considerable amount of money by participating. The text below will tell you how the amount you earn will be determined.

Anonymity

You will never be asked to reveal your identity to anyone during the decision-making part of the experiment. Neither the experimenters nor the other subjects will be able to link you to any of your decisions. In order to keep your decisions private, please do not reveal your choices to any other participant. The following means help to guarantee anonymity:

Non-Computerized Experiment and Private Code

The task you have to complete during the experiment is conducted in private on a printed form; that is, the experiment is not computerized. You have drawn a small sealed envelope from a box upon entering the room. PLEASE DO NOT OPEN YOUR ENVELOPE BEFORE THE EXPERIMENT STARTS. Your envelope contains your participation number. We will refer to it as "your private code" in the following. Your private code is the only identification used during the experiment and you will also need it to collect your cash payments.

When you have completed your task in the experiment you will be asked to write your private code on the front page of your form, to put the form in a new (larger) envelope, to seal the envelope, and to put it in a box located at the front door of the room you are sitting in. It is important that you do not write anything on the envelope, it should be left blank. It is also important that you keep the card with your private code: you need it to collect your earnings!

Cash Payments

Cash payments can be collected from tomorrow onwards in **room w.4.36** in the fourth floor (South/West) of this building. You will present your private code to an admin staff person (Mr. ...) and you will receive your cash payment in exchange. The admin staff person will not know who has done what and why, nor how payments were generated. No experimenter will be present in the room when you collect your money. Also, the private codes of this experiment will be mixed up with the codes of other experiments. This will again help to guarantee that the amount you earn cannot be linked to your decisions. Mr. ... is available from Monday to Friday between 9 a.m. and noon and between 2 p.m. and 3 p.m. in room w.4.36 in the fourth floor (South/West) of this building. Please collect your earnings within a weak. [You find those details also on the card displaying your private code.]

Detailed Instructions

No Talking Allowed

Please read this document carefully and do not talk to any other participant until the experiment is over. If there is anything that you don't understand, please raise your hand. An experimenter will approach you and clarify your questions in private. In about ten minutes this document (the front page included) will also be read aloud (by an experimenter).

Two Groups and Two Different Tasks

Before the experiment starts, the participants in this room will be randomly divided into two groups of equal size (see the text on the next page for details). The groups are called **Group A** and **Group B.** Members of Group A will be seated in this room, members of Group B will be seated in the adjacent laboratory. Each **member of Group A** will be asked to **make** a series of **ten decisions** that affect not only her or his own earnings but also the earnings of a member of Group B. The **members of Group B do not have a decision to make** in this experiment - their earnings will depend on the decisions of Group A members alone. Members of group B will be asked to fill out a questionnaire. This is their only task in this experiment.

Matching

After randomly assigning roles (member of Group A, or member of Group B) to participants, each member of Group A is anonymously paired with a member in Group B**.** The matching is 1:1; that is, each member of Group A is exactly matched with one member of Group B and vice versa. You will **never learn the identity of the member of the other group you are paired with.** In the same way, the member of the other group you are paired with will not learn your identity. In the following we call the member of the other group you are matched with **the other person.**

Task of Members of Group A

If you become a member of Group A you will be asked to make **ten decisions. In each of the ten decision problems** you are asked to decide between **two alternatives** which are called **LEFT** and **RIGHT.** Each alternative implies earnings for you and the other person. The ten decision problems will be presented as rows in a table. Note that only one of the ten decisions will be taken into consideration for the payoff determination - more on this below. Each decision problem will look like this:

LEFT		**Your Choice**	**RIGHT**	
you receive	**other person receives**		**you receive**	**other person receives**
a Euros	*b* Euros	LEFT ◯ ◯ RIGHT	*c* Euros	*d* Euros

The lower case letters in the cells of the decision problem displayed here are for illustration only, in the experiment the letters will be replaced by numbers. If you have been assigned the role of a member of Group A, if in this particular decision problem you choose LEFT, and if this particular problem is chosen as the payoff relevant one, then you receive earnings of *a* Euros while the other person will receives earnings of *b* Euros. Similarly, if you choose RIGHT, you receive *c* Euros and the other person receives *d* Euros. The table on the last page of this document displays the 10 decision problems each Group A member faces. The form members of Group A will receive will contain exactly two pages, the first page is an empty cover page, the second page contains the table on the last page of the current instructions (and nothing else)!

Task of Members of Group B

If you become a member of Group B you will be asked to fill out a two-page questionnaire. The form members of Group B will receive will contain exactly three pages, the first page is an empty cover page, the other two pages contain the questionnaire.

Show-Up Fee

Each participant in this experiment will receive a show-up fee of 4 Euros for participating. In addition, each participant receives earnings as specified in the next two paragraphs. That is, the **final payoff** of a participant **is the sum of** two parts, the **show-up fee plus** the **earnings in the experiment** (as specified below).

Your Earnings if You Are a Member of Group A

If you become a member of Group A your earnings and the earnings of the other person are determined as follows: At the end of the experiment (after you have made the ten choices in private), one of the 10 decision problems will be randomly selected as the payoff-relevant one. For this purpose an experimenter with a bingo cage containing ten balls numbered 1-10 will go from one member of Group A to the next starting on the leftmost cubicle of the first row. Please make sure that your completed form is closed when the experimenter approaches you. The experimenter will ask you to draw one of the balls with the device designated for that purpose. **The number on the ball gives the decision task that will be used to determine your earning and that of the other person.** Your actual earnings and those of the other person correspond exactly to the payoffs in the alternative (LEFT or RIGHT) you have chosen in that specific decision problem. You will be asked by the experimenter to write the number of the payoff-relevant decision problem on the cover page of your form. You (but no one else) will then be given the opportunity to take in private a look at your choice in the payoff-relevant decision problem. Then you will be asked to label (in private) the cover sheet of the form with your private code and to seal the form in the envelope.

Your Earnings if You Are a Member of Group B

In addition to the 4 Euros show-up fee each member of Group B will receive the earnings as described in the previous paragraph from exactly one member of Group A.

Role Assignment and Start of the Experiment

After the instructions at hand have been red aloud and all questions have been answered you (and all other participants in this room) will be asked to open the sealed envelope you draw from the box when entering this room. The envelope contains a card with your private code. The code ends with a number. If this number is even, you are a member of Group A, if it is odd, you are a member of Group B. Members of Group A are asked to take a seat at one of the computer terminals with sliding walls in this room. Members of Group B will be escorted to the adjacent room and asked to take a seat at one of the computer terminals with sliding walls in that room. In both rooms computers are (and will remain) switched off. An experimenter will then distribute the forms in each room. Members of Group A will receive a form that contains an empty cover page and a page containing the decision tasks displayed on the next page, members of Group B will receive a form that contains an empty cover page and a two-page questionnaire.

The End of the Experiment

After you have completed your task you will be asked to write your private code on the empty cover page of your form. PLEASE WAIT UNTIL YOU ARE ASKED BEFORE WRITING THE CODE ON THE COVER. Then put the form in the envelope and seal it. Upon leaving the room you are asked to put the envelope in the box located near the front door of the room you are sitting in.

The Ten Decision Tasks for Members of Group A

The table below displays the ten decision problems presented to members of Group A. Members of Group A will be asked to mark in each row whether they prefer the alternative on the left hand side (LEFT) or the alternative on the right hand side (RIGHT). They have to decide for ONE of the two alternatives in each of the ten rows.

The table below is for illustration only. After the role assignment, members of Group A will receive a form that contains two pages, an empty cover page and a page containing exactly the table below (and nothing else).

Dec. Nr.	LEFT		Your Choice	RIGHT	
	you receive	other person receives		you receive	other person receives
1	**8 Euros**	**13 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**
2	**9 Euros**	**13 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**
3	**10 Euros**	**13 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**
4	**11 Euros**	**13 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**
5	**12 Euros**	**13 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**

Dec. Nr.	LEFT		Your Choice	RIGHT	
	you receive	other person receives		you receive	other person receives
6	**8 Euros**	**7 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**
7	**9 Euros**	**7 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**
8	**10 Euros**	**7 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**
9	**11 Euros**	**7 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**
10	**12 Euros**	**7 Euros**	LEFT ◯ ◯ RIGHT	**10 Euros**	**10 Euros**

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2013-25 **Rudolf Kerschbamer:** The geometry of distributional preferences and a non-parametric identification approach

2013-24 **Nadja Klein, Michel Denuit, Stefan Lang, Thomas Kneib:** Nonlife ratemaking and risk management with bayesian additive models for location, scale and shape

2013-23 **Nadja Klein, Thomas Kneib, Stefan Lang:** Bayesian structured additive distributional regression

2013-22 **David Plavcan, Georg J. Mayr, Achim Zeileis:** Automatic and probabilistic foehn diagnosis with a statistical mixture model

2013-21 **Jakob W. Messner, Georg J. Mayr, Achim Zeileis, Daniel S. Wilks:** Extending extended logistic regression to effectively utilize the ensemble spread

2013-20 **Michael Greinecker, Konrad Podczeck:** Liapounoff's vector measure theorem in Banach spaces *forthcoming in Economic Theory Bulletin*

2013-19 **Florian Lindner:** Decision time and steps of reasoning in a competitive market entry game

2013-18 **Michael Greinecker, Konrad Podczeck:** Purification and independence

2013-17 **Loukas Balafoutas, Rudolf Kerschbamer, Martin Kocher, Matthias Sutter:** Revealed distributional preferences: Individuals vs. teams

2013-16 **Simone Gobien, Björn Vollan:** Playing with the social network: Social cohesion in resettled and non-resettled communities in Cambodia

2013-15 **Björn Vollan, Sebastian Prediger, Markus Frölich:** Co-managing common pool resources: Do formal rules have to be adapted to traditional ecological norms?

2013-14 **Björn Vollan, Yexin Zhou, Andreas Landmann, Biliang Hu, Carsten Herrmann-Pillath:** Cooperation under democracy and authoritarian norms

2013-13 **Florian Lindner, Matthias Sutter:** Level-k reasoning and time pressure in the 11-20 money request game *forthcoming in Economics Letters*

2013-12 **Nadja Klein, Thomas Kneib, Stefan Lang:** Bayesian generalized additive models for location, scale and shape for zero-inflated and overdispersed count data

2013-11 **Thomas Stöckl:** Price efficiency and trading behavior in limit order markets with competing insiders *forthcoming in Experimental Economics*

2013-10 **Sebastian Prediger, Björn Vollan, Benedikt Herrmann:** Resource scarcity, spite and cooperation

2013-09 **Andreas Exenberger, Simon Hartmann:** How does institutional change coincide with changes in the quality of life? An exemplary case study

2013-08 **E. Glenn Dutcher, Loukas Balafoutas, Florian Lindner, Dmitry Ryvkin, Matthias Sutter:** Strive to be first or avoid being last: An experiment on relative performance incentives.

2013-07 **Daniela Glätzle-Rützler, Matthias Sutter, Achim Zeileis:** No myopic loss aversion in adolescents? An experimental note

2013-06 **Conrad Kobel, Engelbert Theurl:** Hospital specialisation within a DRG-Framework: The Austrian case

2013-05 **Martin Halla, Mario Lackner, Johann Scharler:** Does the welfare state destroy the family? Evidence from OECD member countries

2013-04 **Thomas Stöckl, Jürgen Huber, Michael Kirchler, Florian Lindner:** Hot hand belief and gambler's fallacy in teams: Evidence from investment experiments

2013-03 **Wolfgang Luhan, Johann Scharler:** Monetary policy, inflation illusion and the Taylor principle: An experimental study

2013-02 **Esther Blanco, Maria Claudia Lopez, James M. Walker:** Tensions between the resource damage and the private benefits of appropriation in the commons

2013-01 **Jakob W. Messner, Achim Zeileis, Jochen Broecker, Georg J. Mayr:** Improved probabilistic wind power forecasts with an inverse power curve transformation and censored regression

2012-27 **Achim Zeileis, Nikolaus Umlauf, Friedrich Leisch:** Flexible generation of e-learning exams in R: Moodle quizzes, OLAT assessments, and beyond

2012-26 **Francisco Campos-Ortiz, Louis Putterman, T.K. Ahn, Loukas Balafoutas, Mongoljin Batsaikhan, Matthias Sutter:** Security of property as a public good: Institutions, socio-political environment and experimental behavior in five countries

2012-25 **Esther Blanco, Maria Claudia Lopez, James M. Walker:** Appropriation in the commons: variations in the opportunity costs of conservation

2012-24 **Edgar C. Merkle, Jinyan Fan, Achim Zeileis:** Testing for measurement invariance with respect to an ordinal variable *forthcoming in Psychometrika*

2012-23 **Lukas Schrott, Martin Gächter, Engelbert Theurl:** Regional development in advanced countries: A within-country application of the Human Development Index for Austria

2012-22 **Glenn Dutcher, Krista Jabs Saral:** Does team telecommuting affect productivity? An experiment

2012-21 **Thomas Windberger, Jesus Crespo Cuaresma, Janette Walde:** Dirty floating and monetary independence in Central and Eastern Europe - The role of structural breaks

2012-20 **Martin Wagner, Achim Zeileis:** Heterogeneity of regional growth in the European Union

2012-19 **Natalia Montinari, Antonio Nicolo, Regine Oexl:** Mediocrity and induced reciprocity

2012-18 **Esther Blanco, Javier Lozano:** Evolutionary success and failure of wildlife conservancy programs

2012-17 **Ronald Peeters, Marc Vorsatz, Markus Walzl:** Beliefs and truth-telling: A laboratory experiment

2012-16 **Alexander Sebald, Markus Walzl:** Optimal contracts based on subjective evaluations and reciprocity

2012-15 **Alexander Sebald, Markus Walzl:** Subjective performance evaluations and reciprocity in principal-agent relations

2012-14 **Elisabeth Christen:** Time zones matter: The impact of distance and time zones on services trade

2012-13 **Elisabeth Christen, Joseph Francois, Bernard Hoekman:** CGE modeling of market access in services

2012-12 **Loukas Balafoutas, Nikos Nikiforakis:** Norm enforcement in the city: A natural field experiment *forthcoming in European Economic Review*

2012-11 **Dominik Erharter:** Credence goods markets, distributional preferences and the role of institutions

2012-10 **Nikolaus Umlauf, Daniel Adler, Thomas Kneib, Stefan Lang, Achim Zeileis:** Structured additive regression models: An R interface to BayesX

2012-09 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** History repeating: Spain beats Germany in the EURO 2012 Final

2012-08 **Loukas Balafoutas, Glenn Dutcher, Florian Lindner, Dmitry Ryvkin:** The optimal allocation of prizes in tournaments of heterogeneous agents

2012-07 **Stefan Lang, Nikolaus Umlauf, Peter Wechselberger, Kenneth Harttgen, Thomas Kneib:** Multilevel structured additive regression

2012-06 **Elisabeth Waldmann, Thomas Kneib, Yu Ryan Yu, Stefan Lang:** Bayesian semiparametric additive quantile regression

2012-05 **Eric Mayer, Sebastian Rueth, Johann Scharler:** Government debt, inflation dynamics and the transmission of fiscal policy shocks *forthcoming in Economic Modelling*

2012-04 **Markus Leibrecht, Johann Scharler:** Government size and business cycle volatility; How important are credit constraints? *forthcoming in Economica*

2012-03 **Uwe Dulleck, David Johnston, Rudolf Kerschbamer, Matthias Sutter:** The good, the bad and the naive: Do fair prices signal good types or do they induce good behaviour?

2012-02 **Martin G. Kocher, Wolfgang J. Luhan, Matthias Sutter:** Testing a forgotten aspect of Akerlof's gift exchange hypothesis: Relational contracts with individual and uniform wages

2012-01 **Loukas Balafoutas, Florian Lindner, Matthias Sutter:** Sabotage in tournaments: Evidence from a natural experiment *published in Kyklos*

University of Innsbruck

Working Papers in Economics and Statistics

2013-25

Rudolf Kerschbamer

The geometry of distributional preferences and a non-parametric identification approach

Abstract
This paper proposes a geometric delineation of distributional preference types and a non-parametric approach for their identification in a two-person context. It starts with a small set of assumptions on preferences and shows that this set (i) naturally results in a taxonomy of distributional archetypes that nests all empirically relevant types considered in previous work in economics and social psychology; and (ii) gives rise to a clean experimental test design that discriminates between archetypes according to core features of preferences rather than properties of specific modelling variants. As a by-product the test yields a two-dimensional index of preference intensity.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)