RATIONALITY & COMPETITION
CRC TRR 190

Welfare-Based Altruism

Yves Breitmoser (HU Berlin)
Pauline Vorjohann (HU Berlin)

Discussion Paper No. 89

March 28, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Welfare-based altruism

Yves Breitmoser[*]
Humboldt University Berlin

Pauline Vorjohann
Humboldt University Berlin

March 27, 2018

Abstract

Why do people give when asked, but prefer not to be asked, and even take when possible? We show that standard behavioral axioms including separability, narrow bracketing, and scaling invariance predict these seemingly inconsistent observations. Specifically, these axioms imply that interdependence of preferences ("altruism") results from concerns for the welfare of others, as opposed to their mere payoffs, where individual welfares are captured by the reference-dependent value functions known from prospect theory. The resulting preferences are non-convex, which captures giving, sorting, and taking directly. Re-analyzing choices of 981 subjects in 83 treatments covering many variants of dictator games, we find that individual reference points are distributed consistently across studies, allowing us to classify subjects as either non-givers, altruistic givers, or social pressure givers and use welfare-based altruism to reliably predict giving, sorting, and taking across experiments.

JEL codes: C91, D64, D03
Keywords: Social preferences, axiomatic foundation, robustness, giving, charitable donations

[*] We thank Friedel Bolle, Tore Ellingsen, Ernst Fehr, Urs Fischbacher, Susanne Goldlücke, Steffen Huck, Dorothea Kübler, Matthias Lang, Sebastian Schweighofer-Kodritsch, Justin Valasek, Georg Weizsäcker, seminar participants in Berlin, Konstanz and Cologne as well as audiences in Munich (CESifo area conference), Barcelona (IMEBESS 2017) and Lisbon (EEA/ESEM 2017) for many helpful comments. Financial support of the DFG (project BR 4648/1 and CRC TRR 190) is greatly appreciated. Corresponding author: Yves Breitmoser. Address: Spandauer Str. 1, 10099 Berlin, Germany, email: yves.breitmoser@hu-berlin.de, Telephone: +49 30 2093 99408.

1 Introduction

Altruism is widely defined as a concern for the well-being of others. This definition appears to be self-explanatory and to lend itself easily to economic modeling. Yet, any attempt at representing altruistic preferences by means of a utility function seems to prove the opposite. In a seminal paper, Andreoni and Miller (2002) showed that giving in the dictator game is well-captured by simple CES preferences—over the payoff pair of "dictator" and "recipient"—while subsequent research showed that no such utility function is compatible with giving in extensions of the basic dictator game that more realistically capture social interactions outside laboratories. For example, if we allow the recipient to have income of her own, giving is crowded out only imperfectly (Bolton and Katok, 1998), suggesting that warm glow may additionally affect giving (Korenok et al., 2013). Furthermore, if we allow the dictator to take from the recipient's endowment (List, 2007; Bardsley, 2008), we observe asymmetries between giving and taking, suggesting that the warm glow of giving is weaker than the cold prickle of taking (Korenok et al., 2014), a result that is incompatible with related evidence from public goods games. Or, if we allow subjects to sort out of playing a dictator game (Dana et al., 2006), around half of them do so, in particular those who otherwise would give much to the recipient (Lazear et al., 2012), suggesting that we may need to distinguish between altruistic givers and social-pressure givers (DellaVigna et al., 2012). Overall, depending on how we extend the clinical dictator game, a different model of giving seems to be required. In turn, we currently cannot say that we have a reliable model of the simplest economic activity, giving, or a reliable representation of the simplest social preference, altruism. Since any economic interaction involves giving, this raises a fundamental question: Do neither giving nor the underlying preferences lend themselves to (rigorous) economic modeling, and if so, what does?

Our paper presents an axiomatic approach to modeling giving that allows us to directly address this potential impossibility. The advantage of an axiomatic approach is that it provides a positive result, characterizing the family of utility functions that represent all forms of interdependent preferences under "plausible" assumptions (i.e., axioms). This clarifies the set of candidate models, many of which may not have been "invented" so far, and avoids the difficulties inherent in constructing a model based on evidence from a selected range of experiments—that the constructed model may be just one of many candidates. We show that in addition to the standard axioms completeness, transitivity, and continuity, three simple assumptions, namely separability, narrow bracketing, and scaling invariance, refine the vast set of candidate models surprisingly concisely to models where altruism represents a concern for the welfare of others. Individuals maximize a weighted mean of "individual welfare functions" equivalent to the reference-dependent value functions known from prospect theory. We use the term "individual welfare function" to refer to an individual's utility in one-person decision problems. Our representation result (Proposition 1) establishes that this individual welfare is the yardstick by which a person evaluates the consequences of her actions on others, which provides a formal foundation for the general understanding of altruism. In the one-player case, our model reduces to prospect-theoretic utilities, implying that it is compatible with the range of evidence on choice under risk. This stands in contrast to current models of altruism, all of which represent altruism as concerns for payoffs rather than welfares.[1]

[1] An early proposal of Becker (1974) treats altruism as a concern for the utility of others, which yields a linear equation system that can be solved to represent altruism again as a concern for payoffs of others. The resulting differences to

The theoretical predictions of welfare-based altruism applied to giving show how the model organizes the seemingly inconsistent behavior in laboratory experiments. We demonstrate this explicitly for the large family of generalized dictator games. A dictator's optimal transfer at an interior solution decreases in her own reference point while it increases in the recipient's reference point. This explains how a reallocation of initial endowments affects the optimal transfer, by shifting the players' reference points, and predicts imperfect crowding out. Another feature central to welfare-based altruism is that the resulting preferences are not convex, as individual welfares are S-shaped. Non-convexity directly explains that allowing the dictator to take from the recipient's initial endowment may result in "preference reversals"; this means that a dictator whose optimal choice in a game without the possibility to take is to transfer a positive amount to the recipient may switch to taking from the recipient once this is allowed (List, 2007; Bardsley, 2008). Relatedly, losses in relation to the reference point loom larger than gains, akin to loss aversion, explaining the asymmetries between giving and taking (Korenok et al., 2014). Welfare-based altruism also predicts the existence of "reluctant sharers", i.e. persons who transfer a positive amount to the recipient in a standard dictator game but choose a costly option to sort out of the game when given the chance (Dana et al., 2006; Lazear et al., 2012). Since the recipient never learns about the game if the dictator sorts out, her reference point is not adjusted to the dictator game environment in this case and her welfare remains neutral. Once the dictator enters the game, the recipient is informed about the scope of the interaction and forms a reference point reflecting her expectations, which inflicts a negative externality on a welfare-based altruist. If the dictator believes the recipient would form high expectations once informed, she is best off sorting out and leaving the recipient uninformed. It is worth noting that these predictions are explicit, i.e. the opposite results are ruled out by welfare-based altruism (in a sense made precise in Proposition 2).

After establishing these theoretical predictions, we comprehensively evaluate whether welfare-based altruism indeed captures behavior across giving, sorting, and taking decisions in a quantitatively useful manner—as opposed to overfitting observations due to the additional degrees of freedom (the reference points). To this end, we rely on data from controlled laboratory experiments, which allow us to test models very directly, but as reviewed below, the phenomena observed in the field are very similar. Further, potential concerns about overfitting apply equally to all models, in particular to all behavioral models generalizing the so-called standard models, regardless of whether they are models of choice, probability weighting, strategic beliefs, learning, or social preferences. Yet, outside the context of choice under risk (Harless et al., 1994; Wilcox, 2008; Hey et al., 2010), analyses testing the robustness and thus the "applicability" of models are rare, especially considering the sizes of the respective literatures.[2] Occasionally, this seems to be taken as a suggestion that behavioral models may lack robustness, in particular models of social preferences, which we seek to directly address by a comprehensive analysis of model adequacy.

We first estimate the distributions of individual reference points in the four types of dictator game experiments: standard games, games with generalized endowments, with taking options, and with sorting options. The estimated distributions are surprisingly consistent and generally we find three clusters resembling the non-givers, altruistic givers, and social-pressure givers observed by

standard models in games of complete information are negligible (Kritikos and Bolle, 2005).

[2] The short list of exceptions that we are aware of comprises analyses of learning (Camerer and Ho, 1999), strategic choice in normal-form games (Camerer et al., 2004), stochastic choice in dictator games (Breitmoser, 2013, 2017), and bargaining preferences (De Bruyn and Bolton, 2008).

DellaVigna et al. (2012) in charitable fundraising. Implicitly, this clarifies how the diversity of types is captured in a formally uniform manner by welfare-based altruism.

Next, we re-analyze behavior across a set of nine well-known laboratory experiments on giving, comprising 83 choice conditions and around 6500 decisions from 981 subjects. We examine robustness of the in-sample fit within each set of conditions ("descriptive adequacy") and, more importantly, we analyze robustness across conditions ("predictive adequacy") by predicting behavior based on observations from one set of conditions in another set of conditions—asking: Is giving indeed inconsistent across conditions? We find that it is not. Predictions improve substantially by allowing for reference dependence in altruism, as implied by welfare-based altruism, and consistently so across all combinations of in- and out-of-sample conditions. As robustness check, we examine two alternative approaches of extending the standard CES model of altruism, by capturing either warm glow and cold prickle, or envy and guilt, which both fail to improve on CES altruism reliably (i.e. out-of-sample). Hence, the theoretically predicted reference dependence substantively adds to the understanding of behavior and the improved model fit is not due to the additional degrees of freedom per se. We conclude that reference dependence, and the implied non-convexity of preferences, appears to be a stable behavioral trait affecting giving across conditions.

Finally, we use our estimates to predict (again out-of-sample) "social appropriateness" of actions in dictator, taking and sorting games as examined by Krupka and Weber (2013). Their results suggest that behavior may be norm-guided rather than payoff or welfare concerned, casting general doubts on the applicability of models (such as ours) proposed in the existing literature. We show that the "social appropriateness" they elicit via coordination games strongly correlates with the Rawlsian notion of social welfare implied by our predictions of each player's individual welfare. That is, we show that social appropriateness seems to have a simple and intuitive Rawlsian foundation in individual welfare—which we interpret to lend further credibility to both, welfare-based altruism and social appropriateness, as dual approaches towards analyzing behavior.

In conjunction, our results suggest that understanding interdependence of preferences as concerns for the welfare of others is indeed an instrumental approach to explain and model giving. Since any economic interaction is a form of bi- or multilateral giving, and welfare-based altruism is based on general axioms not related to unilateral giving, there are a number of theoretical and practical implications. The axiomatic analysis establishes an interdependence of concepts as diverse as prospect theory, narrow bracketing, altruism, social appropriateness, and reference dependence, and it predicts a range of behavioral puzzles that survived for about 20 years of experimental research. This underlines the power of axiomatic analyses also in the context of social preferences. Further, our results imply that decision makers are utilitarists (for recent discussions, see Fleurbaey and Maniquet, 2011, and Piacquadio, 2017) but in a manner that was predicted by Rawls: rational agents "do not take an interest in one another's interests" (Rawls, 1971, p. 13). That is, agents are concerned with the welfare of others in the way that these others would perceive it in one-person decision problems, but they are not concerned with their altruism or envy, for example. This in turn provides a normative argument for "preference laundering" (Goodin, 1986) in behavioral analyses of social welfare, i.e. for the neglect of emotions such as altruism or envy in welfare analyses. Finally, by generalizing prospect-theoretic utility, welfare-based altruism addresses a number of practical concerns in the literature, such as providing a unified framework for measuring robustness and heterogeneity of preferences across populations and decision problems (Falk et al., 2017), providing a normatively founded framework for measuring reference points across interactions, thereby facilitating a solution to the long-lasting debate on whether and when reference points reflect a status quo (Kahneman et al., 1991), expectations (Kőszegi and Rabin, 2006), or others' payoffs (Fehr and Schmidt, 1999), and providing a general framework for structural analyses of charitable giving (DellaVigna et al., 2012; Huck et al., 2015).

The paper is organized as follows. Section 2 summarizes the related literature on giving. Section 3 provides our representation result (Proposition 1) and discusses its relation to the literature. Section 4 analyzes the implications for giving theoretically in relation to the stylized facts reviewed in Section 2 (Propositions 2 and 3). Section 5 evaluates welfare-based altruism econometrically by a range of in-sample and out-of-sample analyses. Section 6 concludes. The appendix contains relegated definitions, proofs, and robustness checks.

2 Related literature: Experimental evidence on giving

The dictator game is widely used to study giving. We are analyzing the class of generalized dictator games under complete information. In each game, there are two players, the dictator and the recipient. Player 1 (dictator) is endowed with B_1 tokens and player 2 (recipient) is endowed with B_2 tokens. Player 1 can choose $p_1 \in P_1 \subset \mathbb{R}$, inducing a payoff of p_1 for herself and a payoff of $p_2(p_1) = t(B_1 + B_2 - p_1)$ for player 2. We refer to $t > 0$ as transfer rate, to $B = B_1 + B_2$ as budget, and to $B_1 - p_1$ as transfer from the dictator to the recipient.

Definition 1 (Generalized dictator game)**.** A generalized dictator game Γ is defined by the tuple $\langle B_1, B_2, P_1, t \rangle$. The following variants will be distinguished.

- *Standard dictator game:* $B_1 > 0$, $B_2 = 0$, $P_1 \subseteq [0, B_1]$
- *Generalized endowments:* $B_1 \geq 0$, $B_2 > 0$, $P_1 \subseteq [0, B_1]$
- *Taking game:* $B_1 \geq 0$, $B_2 > 0$, $P_1 \subseteq [0, B_1 + B_2]$
- *Sorting game:* $B_1 > 0$, $B_2 = 0$, $P_1 \subseteq \{[0, B_1], \tilde{p}_1\}$ where $\tilde{p}_1$ is an outside option for player 1 inducing payoffs $(\tilde{p}_1, 0)$, with $\tilde{p}_1 \leq B_1$, and implying that 2 is not informed about 1's choice or the rules of the game.

Table 1 provides an overview of giving as observed in generalized dictator games. Following the early work of Kahneman et al. (1986) and for example Hoffman et al. (1996), comprehensive analyses of behavior in the standard dictator game are presented in Andreoni and Miller (2002) and Fisman et al. (2007). The average share of the budget transferred by dictators varies between 20% and 30%, there is an accumulation of transfers at zero and at the payoff-equalizing option, and there is considerable heterogeneity in transfers between subjects. Varying budget sets B and transfer rates t, observed transfers to a large extent satisfy the generalized axiom of revealed preference, implying that dictator behavior is consistent with well-behaved preference orderings. As a candidate for a utility function representing these preferences, Andreoni and Miller (2002) proposed the

Table 1: Stylized facts about giving in generalized dictator games

Comparative statics in t	The transfer can be either constant, increasing, or decreasing in the transfer rate.
Taking options reduce giving at the extensive and intensive margin	Holding endowments constant, extending the choice set of the dictator to the taking domain transforms some initial givers into takers and reduces average amounts given.
Incomplete crowding out	Reallocating endowment from the dictator to the recipient while holding the overall budget constant leads to a less than one-to-one reduction in the dictator's transfer.
Reluctant sharers	A substantial share of givers in the standard dictator game choose to sort out of the game when given the opportunity.
Outside option attractiveness	As the outside option becomes less attractive, fewer dictators sort out of the game, where nonsharers sort back in first followed by the least generous sharers and successively more and more generous sharers.

CES model of altruism, which, using the formulation of Cox et al. (2007), is given by

$$u(\pi) = \pi_1^{\beta}/\beta + \alpha\, \pi_2^{\beta}/\beta \qquad \text{(CES altruism)}$$

with $\alpha, \beta \in \mathbb{R}$. Here, α represents the degree of altruism, $\beta = 1$ implies efficiency concerns, $\beta \to 0$ yields Cobb-Douglas utilities, and $\beta \to -\infty$ implies equity concerns (Leontief preferences).

Comparative statics in ***t*** In a meta-analysis of about 100 experiments, Engel (2011) shows that dictators' transfers increase in the transfer rate, i.e. as transfers become more efficient. This has been observed earlier by Andreoni and Miller (2002) but, for example, not by Fisman et al. (2007). The individual level analyses of Andreoni and Miller (2002) and Fisman et al. (2007) suggest that this inconsistency may be due to differences in subject heterogeneity. In both studies, the majority of subjects act consistently with CES altruism and can be weakly categorized into three standard cases of this utility function, namely selfish, perfect substitutes, and Leontief. Perfectly selfish preferences imply no reaction to changes in the transfer rate, but dictators increase transfers after increases of t if they consider the payoffs to be imperfect substitutes ($\beta > 0$), and they decrease transfers if they consider payoffs to be imperfect complements ($\beta < 0$). Differing shares of the three types in the population may therefore yield differences in the comparative statics in the transfer rate.

Taking options reduce giving at the extensive and intensive margin Holding initial endowments constant, convexity of preferences implies that the extension of the dictator's option set to negative transfers does not affect the choice of a dictator unless she chooses the boundary solution of giving nothing in a standard dictator game. This strong prediction is implied by most models of giving, including CES altruism for $\beta < 1$, but falsified by a strand of studies on so-called taking games (List, 2007; Bardsley, 2008). Both List and Bardsley found that introducing options to take reduces the share of dictators who give positive amounts, though not always significantly. Furthermore, it reduces average amounts given by those who do give positive amounts, and leads to substantive accumulation at the most selfish option. Korenok et al. (2014) confirm these results. List (2007) and Cappelen et al. (2013b) obtain related results on real-effort versions of taking games. List (2007) and Bardsley (2008) interpret the observed patterns in taking games as an indication that choice is menu dependent and, for example, Korenok et al. (2014) argue that taking might induce cold prickle in the sense of Andreoni (1995). Note that we in contrast argue that the initial assumption of convexity may be violated, as known, for example, from choice under risk.

Incomplete crowding out Reference independence of social preferences, as in CES altruism, implies the so-called *crowding out hypothesis* (Bolton and Katok, 1998): lump-sum transfers from dictator to recipient result in a dollar-for-dollar reduction in voluntary giving. The experimental results on dictator games with generalized endowments unanimously falsify this prediction. In both lab and field experiments, dictators reduce their transfers in response to reallocations of endowments to the recipient, but the observed reduction is significantly lower than predicted, a phenomenon referred to as incomplete crowding out (Bolton and Katok, 1998; Eckel et al., 2005; Korenok et al., 2012, 2013). These findings extend to the domain of taking games (Korenok et al., 2014) and to interactions where the budgets are not windfall but generated through either investment games or real effort tasks (Konow, 2000; Cappelen et al., 2007, 2010, 2013a; Almås et al., 2010; Ruffle, 1998; Oxoby and Spraggon, 2008; Jakiela, 2011, 2015). The evidence on dictator games with endowments generated in real effort tasks further suggests that the origin of initital endowments affects dictator behavior. Compared to a standard dictator game with windfall budget, the change to real effort budgets earned by the dictators themselves leads to a drastic reduction in the proportion of nonzero transfers (Cherry, 2001; Cherry et al., 2002; Cherry and Shogren, 2008; Oxoby and Spraggon, 2008; Jakiela, 2011, 2015; Hoffman et al., 1994). Cappelen et al. (2007) relate the observed endowment effects to social norms and, for example, Korenok et al. (2013) interpret the endowment effects as a sign that warm glow in the sense of Andreoni (1995) affects giving. Outside the literature on social preferences, endowment effects are mostly related to reference dependence of preferences (Kahneman et al., 1991; Tversky and Kahneman, 1991), which in turn will be predicted by our representation result.

Reluctant sharers & outside option attractiveness Turning to sorting games, convexity of preferences implies that a dictator cannot be strictly better off by opting out than by staying in. For, the dictator game offers a budget that is at least as high as the outside option. Convexity also implies that no dictator who transfers a positive amount in the dictator game will opt out, since for such a dictator the outside option must be strictly worse than the allocation she chose in the dictator game. Falsifying this prediction, Dana et al. (2006), Broberg et al. (2007), and Lazear et al. (2012) find that a substantive share ($20-60\%$) of their subjects in sorting games can be classified as reluctant sharers, i.e. as dictators who transfer a positive amount in the standard dictator game but given the opportunity rather opt out. As a result, the average amount shared significantly decreases when a sorting option is added to the standard dictator game. Lazear et al. (2012) also find that (i) making the outside option less attractive while holding the dictator game budget constant does reduce the number of dictators who opt out, but (ii) it also reduces the average amount shared. For, mostly nonsharers and reluctant sharers who share less generously in the dictator game reenter first when opting out becomes less attractive. DellaVigna et al. (2012) and Andreoni et al. (2017)

obtain similar results in field experiments on charitable giving. Related to that, Cappelen et al. (2017) observe a close interaction between the information the recipient receives about the origin of her payment and the transfers made by the dictators (in standard dictator games).

There are again multiple proposals for capturing sorting theoretically. DellaVigna et al. (2012) suggest to model it by allowing for an aversion to "saying no" when asked about donations, which however does not capture the comparative statics observed by Lazear et al. (2012), while for example Andreoni and Bernheim (2009) and Ariely et al. (2009) propose to capture reluctancy by including image concerns. As indicated above, the falsified predictions are closely related to convexity, implying that non-convexity directly predicts reluctancy and sorting decisions.

Other extensions Other interesting variations of the standard dictator game include for example the usage of double blind procedures (e.g. Hoffman et al., 1996), extensions to risky environments (e.g. Krawczyk and Le Lec, 2010; Brock et al., 2013), and variations in the transparency of the relationship between dictator choices and outcomes (e.g. Dana et al., 2007). We do not discuss those in more detail here, as these studies have not been designed to primarily study the shape of social preferences, the scope of the present paper, but rather to study the shape of preferences in relation to uncertainty and transparency.

3 Payoff-based and welfare-based altruism: Foundation

Which family of utility functions represents the possible range of altruistic preferences under general behavioral assumptions? Or put differently, which utility functions may be considered plausible candidates in the first place? Despite the plethora of studies on altruism, we are not aware of one addressing or answering these elementary questions. The axiomatic foundations of inequity aversion, e.g. Rohde (2010) and Saito (2013), provide insightful foundations for the widely-used model of Fehr and Schmidt (1999), but they explicitly use inequity-aversion axioms to establish this particular model and thus do not clarify the set of candidate models in general. In this section, we derive two families of utility functions that represent altruism under behavioral assumptions that are comparably well-accepted in related work on preference foundations, in particular on choice under risk. One family captures payoff-based (CES) altruism and the other one captures welfare-based altruism (generalizing Prospect theory). The only difference in the behavioral foundation lies in an assumption clarifying whether subjects factor out background income (narrow bracketing) or not (broad bracketing).

3.1 Theoretical framework

Decision maker DM has to choose an option $x \in X$ where X is a convex subset of $\mathbb{R}^n$. Each option induces an n-dimensional outcome vector captured by $\pi : X \to \mathbb{R}^n$, with $n \geq 3$.[3] We will refer to π as a payoff function, but nothing in our theoretical analysis is specific to preferences over payoff

[3] Assuming the outcome vector has at least three dimensions simplifies some of the statements made below regarding existence of an additively separable utility representation. It is not crucial for the main result. If there was only one essential dimension, existence of an additively separable representation would be trivial, and if there were exactly two essential dimensions in the outcome vector, then existence of an additively separable representation would be ensured by additionally assuming the hexagon condition of Wakker (1989, p. 47).

profiles. For reasons clarified soon, we also say that π defines the "context" of the decision. We use Π to denote the set of payoff functions (and thus contexts) for which the behavioral assumptions are known to hold true. The image of π is $\pi[X] = \{\pi(x) | x \in X\}$.

Preferences over options $x \in X$ are expressed by $\succsim_\pi$, with $\pi(x) \succsim_\pi \pi(y)$ indicating that $\pi(x)$ is weakly preferred to $\pi(y)$. As usual, the strict preference $\pi(x) \succ_\pi \pi(y)$ indicates $\pi(x) \succsim_\pi \pi(y)$ and $\pi(y) \not\succsim_\pi \pi(x)$. Note that this notation of preference is not redundant. It formally clarifies that DM's preferences are **weakly outcome based**, rather than option based, i.e. that DM is indifferent between two options whenever they induce the same outcomes contingent on context π.[4] Intuitively, this assumption seems to be a concern if outcome-equivalent options with different "connotations" are available, but we consider these options not to be outcome equivalent. Note that we do not assume outcome basedness in the strict sense that preferences are independent of context. Indeed, we specifically allow the preference relation $\succsim_\pi$ to be context dependent. Thus, most preference models discussed in analyses of giving can be represented without violating weak outcome basedness in our sense: warm glow and cold prickle, envy and guilt, menu dependence, image concerns, and of course (CES) altruism. Further, all axiomatic characterizations of preferences that we are aware of imply weak outcome basedness by making the stronger assumption of monotonicity. The direct assumption of weakly outcome-based preferences rather than monotonicity allows us to capture the possibility of non-monotonic outcome-based preferences (which relates, for example, to inequity aversion as the best-known non-monotonic preference).

Given this notation, we impose the following assumptions.[5]

Assumption 1 (Framework**).**

1. *Translatability:* $\pi, \pi' \in \Pi \;\Leftrightarrow\;$ there exists $c \in \mathbb{R}^n$ such that $\pi' = c + \pi$
2. *Outcome image is a cone:* $\pi[X]$ is a cone in $\mathbb{R}^n$, i.e. for all $x \in X$ and all $\lambda \in (0,1)$, there exists $x_\lambda \in X$ such that $\pi(x_\lambda) = \lambda\pi(x)$
3. *Essentialness:* All $n \geq 3$ dimensions are essential, i.e. for all $i \leq n$ and each $\pi \in \Pi$, there exist $p, p' \in \pi[X]$ such that $p \succ_\pi p'$ with $p_{-i} = p'_{-i}$.

First, different payoff functions π and π' differ only by translation, i.e. by addition of constants to all outcome vectors, and in turn, all translations are possible. We refer to these additive constants as the "background income" vector, which in turn characterizes the aforementioned "context" of the decision problem. Distinguishing such contexts is novel in relation to the literature and will allow us to state assumptions about responses to changes in background income or concurrent tasks, as discussed below. Second, the image of the set of options in the outcome space is a cone, i.e. we can think of X as a budget set of a consumer facing linear prices: for any option $x \in X$, an option x_λ, for any $\lambda \in (0,1)$, is available satisfying $\pi(x_\lambda) = \lambda\pi(x)$. This assumption implies that the set of options is rich, in the sense that the set of possible outcomes $\pi[X]$ has positive volume in $\mathbb{R}^n$, which is required for uniqueness of the utility representation. Finally, essentialness

[4] Since we focus on distributive preferences, we neglect so-called belief-based components of preferences, i.e. preferences that depend on one's belief about others' actions. Generalizations allowing for belief-based parameters or references points are straightforward to conceive.

[5] Slightly abusing notation, we identify all $c \in \mathbb{R}^n$ with constant functions so the addition of functions and constants is well defined, i.e. for all $\pi, \pi' \in \Pi$, if $\pi' = \pi + c$ then $\pi'(x) = \pi(x) + c$ for all $x \in X$.

requires that there are no redundant dimensions of the outcome vector from DM's perspective, i.e. DM does not ignore any of the dimensions, which is a necessary condition for uniqueness of the utility representation in all dimensions as well.

3.2 Axiomatic foundation of payoff-based and welfare-based altruism

We analyze the interplay of six axioms. The first two require that $\succsim_\pi$ is a continuous weak order, implying that it can be represented by a utility function. Separability (Axiom 3) ensures that an additively separable utility representation exists: if two options are equivalent in any dimension, then changing the value in this dimension equally for both options does not affect the preference ordering between these options. Axioms 4–6 jointly define the functional form. Here, we will analyze the implications of scaling invariance (Axiom 4) in conjunction with either broad bracketing (Axiom 5) or narrow bracketing (Axiom 6).

Assumption 2 (Axioms)**.** For all $\pi \in \Pi$ and all $x, y \in X$:

1. *Weak order:* $\succsim_\pi$ is complete and transitive.
2. *Continuity:* If $\pi(x) \succ_\pi \pi(y)$, there exists $\varepsilon > 0$ such that $\pi(x') \succ_\pi \pi(y')$ for all $x' : d(x', x) < \varepsilon$ and all $y' : d(y', y) < \varepsilon$.
3. *Separability:* For any $x', y' \in X$ such that $\pi_{-i}(x) = \pi_{-i}(x')$ and $\pi_{-i}(y) = \pi_{-i}(y')$, as well as $\pi_i(x) = \pi_i(y)$ and $\pi_i(x') = \pi_i(y')$, we have $\pi(x) \succsim_\pi \pi(y)$ iff $\pi(x') \succsim_\pi \pi(y')$.
4. *Scaling invariance:* There exists a scaling-invariant context $\pi^0 \in \Pi$, i.e. for any $\lambda \in \mathbb{R} : \lambda > 0$, if $\pi^0(x) = \lambda \pi^0(x')$ and $\pi^0(y) = \lambda \pi^0(y')$, then $\pi^0(x) \succsim_{\pi^0} \pi^0(y) \Leftrightarrow \pi^0(x') \succsim_{\pi^0} \pi^0(y')$.
5. *Broad bracketing:* For any $\pi' \in \Pi$, if $\pi(x) = \pi'(x')$ and $\pi(y) = \pi'(y')$, then $\pi(x) \succsim_\pi \pi(y)$ implies $\pi'(x') \succsim_{\pi'} \pi'(y')$.
6. *Narrow bracketing:* For all $c \in \mathbb{R}^n$, $\pi(x) \succsim_\pi \pi(y)$ implies $(\pi + c)(x) \succsim_{\pi+c} (\pi + c)(y)$.

Separability is also known as "independence of equal coordinates" (Wakker, 1989, p. 30). It implies additive separability of the utility function and relates to a broad range of standard assumptions: independence axioms in choice under risk (Wakker and Zank, 2002) or choice under uncertainty (Skiadas, 2013), "independence of irrelevant alternatives" in stochastic choice (Luce, 1959), and separability in social welfare functions (Piacquadio, 2017). Further, additive separability obtains in most utility representations discussed in the literature on altruistic giving, such as CES altruism (Andreoni and Miller, 2002), efficiency concerns (Charness and Rabin, 2002), and impure altruism (Andreoni, 1990; Korenok et al., 2013).[6]

Scaling invariance requires that DM's preferences over two options are robust to scaling the outcome vectors associated with these options. It implies that the utility function is homothetic, which again is satisfied by a broad range of utility functions discussed in the behavioral literature, including CES altruism, inequity aversion, prospect theoretical utilities, and nested CES functions. Scaling invariance is further supported by neuro-physiological evidence showing that the neural

[6] Violations of separability typically capture inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000).

firing rate adapts to the scale of the choice problem (Padoa-Schioppa and Rustichini, 2014)[7] and a host of meta analyses showing that scaling differences between experiments are indeed choice-irrelevant overall. This applies in dictator games (Engel, 2011), ultimatum games (Oosterbeek et al., 2004; Cooper and Dutcher, 2011), trust games (Johnson and Mislin, 2011), and choice under risk (Wilcox, 2008, 2011, 2015).

Finally, broad and narrow bracketing describe behavior in response to changes in "context", which in our case are changes in background income $c \in \mathbb{R}^n$. Broad bracketing assumes that background income is fully factored in when decisions are made, while narrow bracketing assumes that background income is factored out. There is fairly strong evidence, from two strands of literature, that background income is indeed factored out. On the one hand, behavior was shown to be independent of socio-economic background variables such as income or wealth, see e.g. Gächter et al. (2004) and Bellemare et al. (2008, 2011), and in general (Easterlin, 2001). In conjunction with the wide range of results supporting narrow bracketing more generally (e.g. Read et al., 1999, Rabin and Weizsäcker, 2009, Simonsohn and Gino, 2013), this suggests that narrow bracketing is a substantially more adequate behavioral assumption than broad bracketing.

On the other hand, adaptive coding describes the neuro-economic observation that the neuronal representation of subjective values ("utilities") adapts to the range of values in any environment. This enables efficient adaptation to choice environments subject to the physical limitations in neuronal firing rates, and was first observed by Tremblay and Schultz (1999) and subsequently confirmed in a wide range of studies reviewed for example in Padoa-Schioppa and Rustichini (2014) and Camerer et al. (2017). Specifically, Padoa-Schioppa (2009) showed that the baseline activity of the cell encoding the value of a given object generally represents the minimum of the value range, and the upper bound of the activity range of this "value cell" represents the upper bound of the value range. Implicitly, both the scale of the value range and background utility is factored out, yielding choice that satisfies scaling invariance and narrow bracketing simply as a result of the physical limitations in neuronal firing.

As usual, we say that a preference relation $\succsim_\pi$ is represented by a utility function $u_\pi : X \to \mathbb{R}$ if $\pi(x) \succsim_\pi \pi(y) \Leftrightarrow u_\pi(x) \geq u_\pi(y)$ for all $x, y \in X$. Proposition 1 establishes that, in conjunction with the other axioms, preferences compatible with broad bracketing are represented by CES altruism ("payoff-based altruism") and preferences compatible with narrow bracketing are represented by generalized prospect theoretical preferences ("welfare-based altruism").

[7] The best option always has the maximal firing rate and the worst option always has the minimal firing rate, implying that choice is independent of scale after a transition period where the neural firing rate adapts to the scale of the decision problem. See Camerer et al. (2017) for a recent review of the evidence.

Proposition 1. *Given Assumption 1, there exist* $\alpha \in \mathbb{R}^n$, $\beta \in \mathbb{R}$, $\delta \in \mathbb{R}^n$ *and* $r : \Pi \to \mathbb{R}^n$ *such that for all contexts* $\pi \in \Pi$,

$$\text{Axioms 1,2,3,4,5} \quad \Leftrightarrow \quad \succsim_\pi \text{ is represented by } u_\pi(x') = \sum_{i \leq n} \alpha_i \cdot v_i\left[\pi_i(x')\right],$$

$$\text{Axioms 1,2,3,4,6} \quad \Leftrightarrow \quad \succsim_\pi \text{ is represented by } u_\pi(x') = \sum_{i \leq n} \alpha_i \cdot v_i\left[\pi_i(x') - r_i(\pi)\right],$$

for all $x' \in X$, *where the reference points satisfy* $r(\pi^0 + c) = c$ *for all* $c \in \mathbb{R}^n$ *given the scaling invariant context* π^0, *and the value functions* $v_i : \mathbb{R} \to \mathbb{R}$ *satisfy*

$$v_i(p) \underset{\beta \neq 0}{=} \begin{cases} p^\beta/\beta, & \text{if } p \geq 0 \\ -\delta_i \cdot (-p)^\beta/\beta, & \text{if } p < 0 \end{cases} \qquad \text{and} \qquad v_i(p) \underset{\beta = 0}{=} \log(p).$$

The proof is relegated to the appendix. Existence of a continuous weak order (Axioms 1 and 2) implies that, in each context π, the preference relation $\succsim_\pi$ can be represented by some utility function $u_\pi : X \to \mathbb{R}$. Axiom 3 implies that an additively separable utility representation exists (Wakker, 1989), i.e. given context π, value functions $\{v_{\pi,i} : \mathbb{R} \to \mathbb{R}\}_{i \leq n}$ exist such that u_π with

$$u_\pi(x) = \sum_{i \leq n} v_{\pi,i}\big(\pi_i(x)\big) \tag{1}$$

represents $\succsim_\pi$. Broad bracketing implies that these value functions must be equivalent across contexts. Narrow bracketing requires that context shifts (changes in background income) are factored out, which implies that payoffs must be evaluated in relation to some unknown reference points. As a result, there exists a family of functions $\{v_i : \mathbb{R} \to \mathbb{R}\}_{i \leq n}$ and $r : \Pi \to \mathbb{R}^n$ such that

$$u_\pi(x) = \sum_{i \leq n} v_i\big(\pi_i(x)\big) \qquad\qquad u_\pi(x) = \sum_{i \leq n} v_i\big(\pi_i(x) - r_i(\pi)\big) \tag{2}$$

represent $\succsim_\pi$ for all $\pi \in \Pi$ in the cases of broad bracketing and narrow bracketing, respectively. With narrow bracketing, the utility function is equivalently expressed as

$$u_\pi(x) = \sum_{i \leq n} \alpha_i \cdot \left[r_i(\pi) + v_i\big(\pi_i(x) - r_i(\pi)\big)\right], \tag{3}$$

simply adding the reference points in all dimensions (or any other constant; given separability, the utility function is unique up to positive affine transformation). Formulation (3) may appear more intuitive if the reference points differ from zero.[8] Similarly, by uniqueness up to affine transformation, the weights (α_i) are unique up to scaling. A standard restriction here is to require that (α_i) adds up to 1. Finally, scaling invariance pins down the functional form of v_i. By scaling invariance, we know that, focusing on broad bracketing for simplicity here,

$$u_\pi(x) = \sum_{i \leq n} v_i\big(\pi_i(x)\big) \qquad \text{and} \qquad u_{\lambda\pi}(x) = \sum_{i \leq n} v_i\big(\lambda\pi_i(x)\big)$$

[8] It expresses the idea that meeting one's reference point implies a utility exactly equal to the reference point (in case the value function is the power function in Proposition 1). Thus, for example, an individual being \$10 short of their reference point \$1,000,000 would enjoy a higher utility than an individual being \$10 short of their reference point \$20.

with $\lambda \in (0,1)$ both represent $\succsim_\pi$, and both being additively separable, this implies that they are positive affine transformations of one another. Hence, for all $i \leq n$,

$$v_i\big(\lambda\pi_i(x)\big) = a_i(\lambda) + b(\lambda) \cdot v_i\big(\pi_i(x)\big)$$

for some functions $a_i : \mathbb{R} \to \mathbb{R}$ and $b : \mathbb{R} \to \mathbb{R}_+$. By Assumption 1.2, the value function v_i is defined on an interval of positive length, by Axiom 2 it is continuous, and by 1.3 it is not equal to the constant function, which jointly implies that the unique solutions of this Pexider functional equation (Aczél, 1966) are the power and logarithmic functions defined in Proposition 1.[9]

Two technical points appear worth noting. If there exists $x \in X$ such that $\pi^0(x) = 0$, where π^0 denotes the scaling-invariant context, then $\beta > 0$ obtains by continuity. Further, if we assume monotonicity, the parameters (α_i, δ_i) are guaranteed to be non-negative. While this appears plausible in many cases, it would rule out some phenomena resembling inequity aversion, the defining characteristic of which is that preferences are non-monotonic in the opponents' outcomes.[10]

3.3 Discussion

To summarize, broad bracketing induces a context-independent reference point of zero, yielding the well-known CES model of altruism where payoffs are evaluated in absolute terms. This captures altruism as concern for the payoffs of others. Narrow bracketing implies that payoffs are evaluated in relation to reference points $r_i(\pi) = \pi_i(x) - \pi_i^0(x)$, where π^0 is the scaling invariant context existing by Axiom 4. This implies that altruism is a concern for the S-shaped welfares of others known from prospect theory, i.e. for the (individual) welfares they believe the others would derive from the various outcomes in single-person decision problems. Switching from broad bracketing to narrow bracketing is in this sense equivalent to switching from altruism as a concern for the payoff of others to altruism as a concern for the welfare of others. Next, we briefly discuss this observation in relation to prominent strands of literature.

Contexts and narrow bracketing Following Read et al. (1999), narrow bracketing refers to the phenomenon that concurrent decision problems are treated independently by decision makers, implying that other tasks simply provide a background income (the "context") that is factored out. Using this observation as part of an axiomatic foundation is novel and to be distinguished from translation invariance. Specifically, narrow bracketing operates between contexts (changes in background income) and translation invariance operates within contexts.[11] The distinction is noteworthy, as it is narrow bracketing, rather than translation invariance, that is backed by the behavioral and neuroeconomic evidence cited above. Further, narrow bracketing is compatible

[9] For illustrative purposes, assume v_i is also differentiable and let $a_i = 0$ (which removes the logarithmic solution). That is, $v_i(\lambda\,\pi_i) = b(\lambda) \cdot v_i(\pi_i)$, and after taking logarithms on both sides, we obtain for $\tilde{v}_i = \log v_i$ and $\tilde{b} = \log b$,

$$\tilde{v}_i(\lambda\,\pi_i) = \tilde{b}(\lambda) + \tilde{v}_i(\pi_i) \qquad \Rightarrow \qquad \tilde{v}_i'(\lambda\,\pi_i) \cdot \pi_i = \tilde{b}'(\lambda) \qquad \Rightarrow \qquad \tilde{v}_i'(\pi_i) = \beta/\pi_i$$

after taking the derivative with respect to λ and letting $\lambda = 1$. This differential equation has the solution $\tilde{v}_i(\pi_i) = \beta \log \pi_i + \alpha_i$ and reverting the logarithm we obtain $v_i(\pi_i) = \alpha_i \cdot \pi_i^\beta$.

[10] For example, inequity averse subjects prefer $(10,9)$ over $(11,20)$, or $(0,0)$ over $(1,9)$. Without monotonicity, welfare-based altruism can capture such preferences, and in this way, it can also capture rejections in ultimatum games.

[11] Translation invariance requires that if one pair of options yields payoff vectors π_x and π_y, and another pair of options yields $\pi_x + r$ and $\pi_y + r$ for some $r \in \mathbb{R}$, then the respective choices must be equivalent (see for example Skiadas, 2013). In contrast, narrow bracketing poses no restriction for different pairs of options.

with scaling invariance, as one operates between contexts and the other one operates within contexts, implying that by formalizing narrow bracketing as a preference axiom we can acknowledge both observations on human behavior in the axiomatic analysis. Translation invariance (within contexts) is obviously not compatible with scaling invariance.

Reference dependence Narrow bracketing implies reference dependence; it implies the existence of reference points with the testable prediction that reference points move 1:1 as the background income changes. This yields a foundation of reference dependence without an ex-ante specification of reference points. The result generalizes existing axiomatic foundations of prospect theoretical utilities, which so far explicitly assume existence of a reference point, where the reference point is either an exogenously defined payoff vector (Wakker and Tversky, 1993; Wakker and Zank, 2002) or a well-defined option (Schmidt, 2003). Further, the link between narrow bracketing and reference dependence is established based on axioms not related specifically to altruism or giving, underlining its generality and corroborating the observation that both narrow bracketing and reference dependence build on a wealth of empirical evidence (outside prospect theory, see for example Kőszegi and Rabin, 2007, 2009, for discussion).

Reference points The reference points may reflect any (weighted) mean of status quo (Kahneman et al., 1991) and expectations (Kőszegi and Rabin, 2006), since any such mean satisfies the above condition that adding a fixed vector of background incomes to all payoff profiles raises the vector of reference points by exactly that vector (assuming expectations are independent of the background income). Further, reference points are idiosyncratic, reflecting that different decision makers may well disagree about the status quo or the expectations in a given choice task. In turn, the above axioms do not imply normative restrictions about the absolute values of reference points or that there is a normative or objective definition of the status quo.

Context dependence Our notation allows preferences to depend on the payoff function π, while the option set X is assumed to be constant. In general, X may vary as well, and π may vary by more than changes in background income. We have not imposed assumptions linking preferences across such changes in contexts, implying that preferences may change arbitrarily when say the image $\pi[X]$ changes in dimensions other than the background income. The existing literature analyzes how reference points may depend on say $\pi[X]$, usually in the form of expectation-based reference points. In individual decision making this was proposed by Kőszegi and Rabin (2006) and experimentally observed for example in Falk et al. (2011) and Gill and Prowse (2012). In games such expectation-based models imply reciprocity, as discussed in Rabin (1993), Levine (1998), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006). Implicitly, by allowing for expectation-based reference dependence, our simple model of altruism is compatible with these models of reciprocity, in particular once we recognize that the "context" may be a function of previous moves by other players. Then, preferences may immediately reflect the kindness of such moves as discussed in the literature.

Choice under risk Our analysis relates to studies of preferences in choice under risk, see for example Wakker and Tversky (1993) for an analysis assuming differentiability and Skiadas (2013) for a recent analysis without differentiability. Most axioms in this branch of literature are similar to those imposed above, suggesting the possibility of constructing a general, unified foundation of behavior. The main difference of our analysis to this literature is the formal distinction of contexts. This is substantial, as it allows us to analyze narrow bracketing instead of translation invariance, which in turn is compatible with scaling invariance. The similarities are that analyses of choice under risk also work with existence of a weak order, continuity, and generally an independence assumption yielding additive separability across possible outcomes. Skiadas (2016) shows that system of axioms including scaling invariance implies a form of CES preferences that is similar to CES altruism as characterized above (i.e. not to welfare-based altruism), while one including translation invariance implies exponential rather than power utilities resembling constant absolute risk aversion. His results suggest that scaling invariance and translation invariance are mutually exclusive in axiomatic foundations, although both tend to be confirmed in behavioral meta studies, a conflict that is resolved in our analysis.

Inequity aversion To discuss the relation to inequity aversion (Fehr and Schmidt, 1999), recall that using $|r|_+$ to be equal to r if $r > 0$ and equal to zero otherwise, inequity averse decision makers have utilities

$$u(\pi) = \pi_1 - \alpha_1 \cdot |\pi_1 - \pi_2|_+ - \alpha_2 \cdot |\pi_2 - \pi_1|_+,$$

where guilt has weight α_1 and envy has weight α_2. In relation to welfare-based altruism, inequity aversion relaxes additive separability (in a specific manner) and violates narrow bracketing, while it satisfies scaling invariance and broad bracketing. Yet, welfare-based altruism contains FS inequity aversion as a special case in constant-sum games. To see this, assume that $\pi_1(x) + \pi_2(x)$ is constant for all options $x \in X$ and let (r_1, r_2) denote the reference points. Inequity aversion corresponds to the special case $r_1 = r_2 = (\pi_1 + \pi_2)/2$ of linear welfare-based altruism. Assuming "loss aversion", i.e. $\delta > 1$, this yields a piecewise linear utility function which is steeper in case of envy ($\pi_1 < \pi_2$) than in case of guilt ($\pi_1 \geq \pi_2$), implying inequity aversion as defined by Fehr and Schmidt (1999).

Warm glow and cold prickle Andreoni (1989, 1990) argues that warm glow and cold prickle affect behavior in the sense that players derive utility directly from the amount they transfer or take. In dictator games such preferences can rationalize incomplete crowding-out. Using e_1 and e_2 to denote the players' endowments, warm glow is proportional to the amount $e_1 - \pi_1$ transferred from the own endowment (weight α_1) and cold prickle is proportional to the amount $e_2 - \pi_2$ taken from the other endowment (weight α_2), implying utilities

$$u(\pi) = \pi_1 + \alpha_1 \cdot |e_1 - \pi_1|_+ - \alpha_2 \cdot |e_2 - \pi_2|_+.$$

The difference to models of "pure" altruism such as CES altruism is that the extent of both warm glow and cold prickle is independent of how the respective other party benefits. If preferences exhibit warm glow and cold prickle, then, similar to welfare-based altruism, a scaling invariant context exists and separability obtains, but neither broad nor narrow bracketing applies. Narrow bracketing is violated as the first term is not reference dependent and broad bracketing is violated because the other terms are reference dependent. That is, narrow and broad bracketing are violated in specific ways and in addition the reference points explicitly equate with the endowments rather than arising implicitly. Hence, comparably strong axioms would be required to behaviorally found such “impure” altruism; axioms that seem difficult to motivate based on independent behavioral results. This substantially limits generalizability to the wide range of interactions where endowments and hence warm glow as well as cold prickle are not explicitly defined, such as bargaining games or mini-dictator games.

4 Implications for giving: Theory

We now apply the representation result (Proposition 1) to study giving of welfare-based altruists and how it relates to the observations made in experiments. By context dependence, the reference points r_1, r_2 may be arbitrary functions of the game $\Gamma = \langle B_1, B_2, P_1, t\rangle$, but we know there exist reference points r_1 and r_2 such that the dictator's utility in a generalized dictator game Γ is

$$u_\Gamma(p_1) = \frac{1}{\beta} \times \left\{ \begin{matrix} (p_1 - r_1)^\beta & \text{if } p_1 \geq r_1 \\ -\delta(r_1 - p_1)^\beta & \text{if } p_1 < r_1 \end{matrix} \right\} + \frac{\alpha}{\beta} \times \left\{ \begin{matrix} (t(B - p_1) - r_2)^\beta & \text{if } p_2(p_1) \geq r_2 \\ -\delta(r_2 - t(B - p_1))^\beta & \text{if } p_2(p_1) < r_2 \end{matrix} \right\}.$$

As above, α is the degree of altruism, δ is the degree of loss-aversion, and β captures the trade-off between efficiency and equity concerns; $\frac{1}{1+\beta}$ is the elasticity of substitution between dictator's and recipient's well-being. Without loss of generality, we assume that δ is the same for both players, and for notational simplicity, we skip the limiting case $\beta = 0$.

The reference points contain the players' background incomes as additive constants by Proposition 1, which we represent by the players' minimal payoffs, and may otherwise be arbitrarily complex functions of the game characteristics such as payoff function and option sets. For the following analysis, we express this richness as a two-parametric family of functions of the ranges of payoffs. Specifically, for each player we determine how much of her endowment is at the disposal of the dictator and allow her to form expectations about how the surplus is allocated. Every player expects to be allocated share $w_1 \in [0,1]$ of the surplus she contributes and share $w_2 \in [0,1]$ of the surplus contributed by the other player, on top of $\min p_1$ and $\min p_2$ (the “background incomes”).

Assumption 3. In game $\Gamma = \langle B_1, B_2, P_1, t\rangle$, the reference points satisfy, for some $w_1, w_2 \in [0,1]$,

$$r_1(\Gamma) = \min p_1 + w_1 \cdot (B_1 - \min p_1) + w_2 \cdot (B_2 - \min p_2/t),$$
$$r_2(\Gamma) = \min p_2 + w_2 \cdot t \cdot (B_1 - \min p_1) + w_1 \cdot (t \cdot B_2 - \min p_2).$$

This model contains status-quo-based reference points ($w_1 = w_2 = 0$) and strict expectations-based reference points ($w_1 + w_2 = 1$) as the most notable special cases, and by allowing for $w_1 + w_2 \in (0,1)$ all convex combinations are also included. Below, we even allow for $w_1 + w_2 > 1$, to capture social-pressure givers, but this case is theoretically rather simple (social-pressure givers always choose corner solutions) and will therefore be sidelined initially.

Dictators are welfare-based altruists denoted as $\Delta = (\alpha, \beta, \delta, w_1, w_2)$. Besides satisfiability of reference points ($w_1 + w_2 \leq 1$), we assume that dictators are imperfectly altruistic ($0 \leq \alpha \leq 1$), imperfectly efficiency concerned ($0 < \beta < 1$), and weakly loss averse ($\delta \geq 1$). Both $0 < \beta < 1$ and

$\delta \geq 1$ are standard assumptions in, for example, prospect theoretical analyses, ensuring S-shaped utilities and avoiding loss seeking, which we therefore adopt as well. Weak altruism ($\alpha \leq 1$) is a standard assumption in analyses of social preferences and $\alpha \geq 0$ is assumed without loss of generality as egoism ($\alpha = 0$) is equivalent to spite ($\alpha < 0$) in the games we analyze.

Definition 2. Dictator $\Delta = (\alpha, \beta, \delta, w_1, w_2)$ is called **regular** if she exhibits imperfect altruism ($0 \leq \alpha \leq 1$), weak efficiency concerns ($0 < \beta < 1$), loss aversion ($\delta \geq 1$), and satisfiability ($w_1 + w_2 \leq 1$).

Proposition 2 formally characterizes giving of welfare-based altruists to provide the basic intuition. Our subsequent result will explore the relations to the stylized facts discussed above.

Proposition 2. *There exist thresholds* $(\delta^-(\Gamma), \delta^+(\Gamma))$ *in terms of the degree of loss aversion* δ *such that for almost all regular dictators* $\Delta = (\alpha, \beta, \delta, w_1, w_2)$*,*

$$(p_1^*, p_2^*) = \begin{cases} (p_1^+(\Gamma), p_2^+(\Gamma)), & \textit{if } \delta > \max\{\delta^+(\Gamma), \delta^-(\Gamma)\} \qquad \textbf{\textit{(interior solution)}} \\ (\max p_1, \min p_2), & \textit{if } \delta < \delta^+(\Gamma) \qquad \textbf{\textit{(egoistic solution)}} \\ (\min p_1, \max p_2), & \textit{if } \delta < \delta^-(\Gamma) \qquad \textbf{\textit{(altruistic solution)}} \end{cases}$$

with $(p_1^+(\Gamma), p_2^+(\Gamma))$ *referring to the interior solution*

$$p_1^+(\Gamma) = \frac{tB + c_\alpha r_1(\Gamma) - r_2(\Gamma)/t}{c_\alpha + 1} \qquad\qquad p_2^+(\Gamma) = \frac{t\,c_\alpha (B - r_1(\Gamma)) + r_2(\Gamma)}{c_\alpha + 1}$$

with $c_\alpha := (\alpha t^\beta)^{\frac{1}{1-\beta}}$*. In any generalized dictator game* $\Gamma = \langle B_1, B_2, P_1, t \rangle$*, there are regular dictators choosing the interior solution and regular dictators choosing either the egoistic solution* $(\max p_1, \min p_2)$ *or the altruistic solution* $(\min p_1, \max p_2)$*.*

That is, there are up to three types of welfare-based altruists: some give nothing or take all (choosing the lower bound), some give a bit (choosing an interior solution), and some give all (choosing the upper bound). In the interior solution, both reference points are satisfied, which implies that many possible decisions can be ruled out. Further, the types of welfare-based altruists are defined using simple thresholds $(\delta^-(\Gamma), \delta^+(\Gamma))$ in terms of the degree of loss aversion δ, and after ruling out loss seeking ($\delta < 1$), the cases $\delta < \delta^+(\Gamma)$ and $\delta < \delta^-(\Gamma)$ are mutually exclusive. Thus, δ allows us to rank dictators by their propensity to choose either of the corner solutions. Dictators with a low degree of loss aversion δ have a high propensity to choose a corner solution, evaluating the extra costs of not satisfying a reference point as low, and dictators with a high degree of loss aversion pick an interior solution. The type of corner solution chosen by dictators with low δ depends on both their degree of altruism α and the transfer efficiency t, where the altruistic corner solution is relevant if $\alpha t^\beta \geq 1$ and the egoistic corner solution otherwise. The thresholds are continuous in the game parameters $\langle B_1, B_2, P_1, t \rangle$ and the preference parameters $(\alpha, \beta, w_1, w_2)$. This allows us to characterize the comparative statics of behavior across dictator games. Finally, and most importantly for the implications on giving, the interior solution has very intuitive comparative statics: The recipient's payoff is decreasing in the dictator's reference point r_1, increasing in the recipient's reference point r_2 and budget B, and increasing in the transfer rate t.

In conjunction with the similarly intuitive comparative statics of the behavioral thresholds

$\delta^+(\Gamma)$ and $\delta^-(\Gamma)$, this directly predicts the stylized facts observed in the literature (Table 1). Proposition 3 establishes this formally, a detailed discussion follows. As above, we say a dictator is a "giver" if she transfers some of her endowment to the recipient, she is a "taker" if the net-transfer is negative, and comparing two games, we say that the range of taking options is extended if $B = B_1 + B_2$ is held constant but the maximal dictator transfer $\max p_1$ increases.

Proposition 3. *Assume dictators $\Delta = (\alpha, \beta, \delta, w_1, w_2)$ are randomly distributed in $\mathbb{R}^5$ such that dictator Δ has positive density if and only if dictator Δ is regular. All "stylized facts" are implied.*

1. ***Non-convexity** In all games with $P_1 = [0, B]$, some dictators have non-convex preferences.*

2. ***Taking options reduce giving both at the extensive and intensive margin** Introducing a taking option turns some initial givers into takers and reduces average amounts given.*

3. ***Incomplete crowding out** Reallocating initial endowment from the dictator to the recipient results (in expectation) in a payoff increase for the recipient.*

4. ***Efficiency concerns** The recipient's payoff is weakly increasing in the transfer rate.*

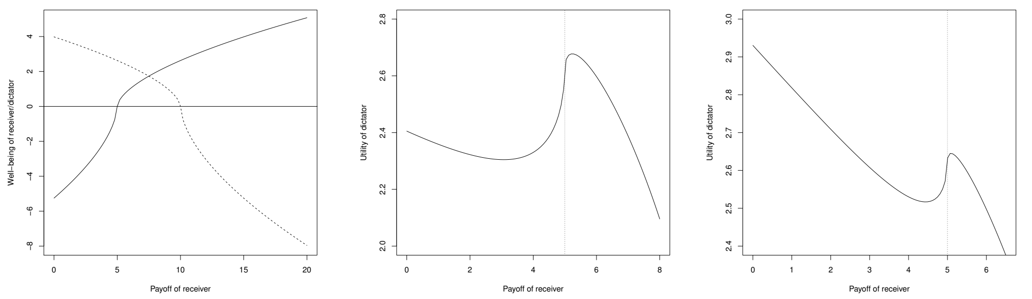
5. ***Reluctant sharers** When an outside option is introduced, some initial givers switch to that option while the behavior of dictators who sort into the game stays unaffected.*

6. ***Social pressure givers** Ceteris paribus, higher susceptibility to social pressure (higher w_2) implies higher transfers in the interior solution but also a higher propensity to choose the outside option in a sorting game.*

Non-convexity of preferences and jumping to take all The most distinctive characteristic of welfare-based altruism is the implied non-convexity of preferences. Figure 1 considers a dictator asked to allocate a budget of 20 tokens between herself and the recipient at a transfer rate of $t = 1$. Suppose that the reference points are $r_2 = 5$ for the recipient and $r_1 = 10$ for the dictator. Figure 1a depicts the trade-off that the dictator faces between her own and the recipient's welfare. The more the dictator allocates to the recipient, the higher is the recipient's welfare (solid curve) but the lower is the dictator's own welfare (dashed curve). The individual welfares are steeper the closer the players are to their respective reference points. For recipient payoffs between 5 and 10, both the recipient and the dictator are in the gain domain, i.e. they achieve payoffs at least as high as their respective reference points, whereas for all other allocations one of them is in the loss domain. Figure 1b depicts the dictator's utility if her weight on the recipient's well-being is $\alpha = 0.3$. This dictator's utility function reaches its maximum at an interior solution where the transfer slightly exceeds the recipient's reference point.

Figure 1c depicts the utility of a slightly less altruistic dictator ($\alpha = 0.2$). This dictator's optimal choice is the corner solution of allocating nothing to the recipient. The S-shaped form of the individual welfare function implies that the deeper the recipient moves into the loss domain, the lower the marginal reduction in recipient welfare for any further token not allocated to him. In conjunction with weak altruism and the correspondingly S-shaped dictator welfare, this implies that dictator utility is not quasi-concave—it bends upwards once the recipient is sufficiently far

Figure 1: Non-convexity of preferences and implications in taking games



(a) Welfares of recipient (solid) and dictator (dashed) with $\beta = 0.6$ and $\delta = 2$

(b) Utility of a dictator with $\alpha = 0.3$

(c) Utility of a dictator with $\alpha = 0.2$

Note: The dictator can choose to allocate x tokens to the recipient, where $x \in [0,20]$. The transfer rate is $t = 1$. The recipient's reference point is $r_2 = 5$ while the dictator's reference point is $r_1 = 10$. The dashed lines in (b) and (c) mark the recipient's reference point.

below his reference point. Ceteris paribus, the lower the weight α that the dictator assigns to the recipient's welfare, the earlier this minimum is reached and the more likely it is that the dictator's utility from choosing the lower bound exceeds her utility in the interior solution.

As a result, dictator behavior is not generally continuous in the game parameters, which in turn predicts the "preference reversals" observed in taking games. To see this, have another look at Figure 1c, now assuming the recipient's reference point equates with his endowment ($B_2 = 5$ and $B_1 = 15$). That is, if the dictator allocates, say, 4 to the recipient, then she actually takes from his endowment. For simplicity, also assume that reference points are invariant to changes in the dictator's choice set, which is covered in the subsequent discussion of "taking games", and let us start with the case that the dictator cannot take from the recipient's endowment. In this case, the dictator cannot implement an allocation with a recipient payoff below his reference point, to the left of the vertical dotted line in Figure 1c, and chooses the interior solution to the right. Now, as we extend the option set by allowing for taking one token from the recipient, allocations to the left of the vertical dotted line become admissible. Initially, upon extending the option set, the dictator's utility at the lower bound is decreasing. The recipient's welfare drops sharply and the dictator is concerned for his welfare. Eventually, upon further extending the option set into the taking domain, the dictator's utility reaches a minimum and starts increasing again. Eventually, the dictator prefers the lower bound to the interior solution and jumps to taking as much as possible. Such a "preference reversal" cannot be observed for the more altruistic dictator in Figure 1b as long as the recipient's payoff is restricted to be non-negative.

Taking games Introducing taking options decreases the recipient's minimal payoff, i.e. his background income. Regardless of whether the recipient has status-quo-based or expectations-based reference points, or a convex combination from the general class in Assumption 3, the recipient's reference point will consequentially decline. The reduction in the recipient's minimal payoff at the same time raises the surplus $B_2 - \min p_2/t$ he contributes, but generically (for all $w_1 < 1$) the first effect dominates. Loosely speaking, the recipient will be happy with less. In turn, the dictator's reference point weakly increases through her partial claim on the increasing surplus contributed by the recipient (if $w_2 > 0$). That is, after introducing taking options, the dictator asks for more. Both effects directly imply, at the intensive margin, that the dictator transfers less in the interior solution, which has the obvious comparative statics in reference points by Proposition 2. In addition, as the lower bound declines, defecting towards the lower bound becomes more attractive for the dictator (recall Figure 1) and with the increase of the own reference point, the interior solution becomes less attractive. As a result, at the extensive margin, dictators are more likely to pick the lower bound, and across the population, the share of regular dictators who choose the lower bound increases while the share of regular dictators who choose the interior solution decreases.

Generalized endowments Assume part of the dictator's endowment is reallocated to the recipient and the dictator cannot take any of it back, i.e. her budget correspondingly declines. Then, the dictator's background income is constant but the surplus she contributes ($B_1 - \min p_1$) decreases, while the recipient's background income increases and his surplus remains constant. As a result, the dictator's reference point declines and the recipient's reference point increases. By the comparative statics of the interior solution, the dictator thus allocates less to herself and more to the recipient at the interior solution, implying incomplete crowding out of endowment reallocations.

Sorting games Lazear et al. (2012) call a dictator a "reluctant sharer" if she transfers a positive amount in a standard dictator game but sorts out when possible. That is, her utility from the interior solution is lower than her utility from the outside option $(\tilde{p}_1, 0)$—assuming the recipient is not informed about the dictator and her options if she sorts out. Remaining uninformed if the dictator sorts out, from the recipient's perspective literally nothing happens, both reference point and payoff are zero, and he remains welfare neutral. This removes the negative externality imposed by the recipient's expectations and may therefore be preferable for the dictator. To see this, assume reference points are just "satisfiable", i.e. $B = r_1 + r_2/t$, and the dictator chooses to satisfy them in the standard dictator game (as opposed to choosing the lower bound). The interior solution generates zero surplus for either player and consequentially zero utility. Then, sorting out is strictly preferable whenever $\tilde{p}_1 > r_1$. If we set $\tilde{p}_1 = B_1$ and start declining it, as in the experiment of Lazear et al. (2012), the condition $\tilde{p}_1 > r_1$ is first violated for dictators with high reference points r_1, who transfer the least at the interior solution. These players are thus predicted to sort in first, regardless of how subjects mix status quo and expectations forming reference points, which corroborates the observation of Lazear et al. (2012) that the least generous dictators sort back in first.

5 Implications for giving: Test on data

We complement the theoretical analysis by testing the model on data from the seminal experiments cited above. We examine whether welfare-based altruism indeed helps improve our understanding of giving in a statistically significant manner. In this way, we can control for the additional degrees of freedom arising in relation to the standard model of payoff-based CES altruism, the two reference points. To be clear, the fact that the theoretical predictions match the comparative statics for all distributions of reference points given "regularity" of dictator preferences strongly suggests that welfare-based altruism does capture giving more reliably than payoff-based altruism—without the

Table 2: The experiments re-analyzed to verify model adequacy

		Abbreviation	#Treatments	# Subjects	#Observations
Dictator games	Andreoni and Miller (2002)	AM02	8	176	1408
	Harrison and Johnson (2006)	HJ06	10	56	560
Generalized endowments					
	Cappelen et al. (2007)	CHST07	11	96	190
	Korenok et al. (2012)	KMR12	8	34	272
	Korenok et al. (2013)	KMR13	18	119	2142
Taking (and generalized endowments)					
	List (2007)	List07	3	120	120
	Bardsley (2008)	Bard08	6	180	180
	Korenok et al. (2014)	KMR14	9	106	954
Sorting	Lazear et al. (2012)	LMW12	8	94	518
Aggregate		Pooled	83	981	6578

necessity of fine-tuning parameters. Yet, in future applications, the additional degrees of freedom may increase the susceptibility to overfitting, as observed for example by Hey et al. (2010) analyzing models of choice under risk. This would limit the model's usefulness when predicting, say, implications of policy interventions or mechanism designs, or simply when interpreting behavior. This is tested next. For the lack of comparable analyses in the existing literature, we include a number of well-known models as benchmarks to provide some context.

5.1 The data

Table 2 summarizes the types of dictator games and data sets we re-analyze. All of them represent seminal papers run for the purpose of characterizing preferences underlying giving, rendering them adequate also for our purpose of validating utility representations of preferences. In total, we analyze behavior across 9 experiments, 83 treatments and 6500 observations. In relation to comparable studies of model validity, e.g. on choice under risk, this represents a very comprehensive data set, promising reliable results.

To our knowledge, our data set includes all experiments on generalized dictator games, i.e. with generalized endowments, taking, or sorting options, complete information, at least three treatments, manual entry of choices, and freely available data sets. The focus on experiments with at least three treatments facilitates statistically informative likelihood ratios but it precludes small experiments, most notably a seminal paper on sorting (Dana et al., 2006). The focus on games with complete information facilitates a unified theoretical treatment but precludes field experiments on charitable giving (such as DellaVigna et al., 2012) and experiments on moral wiggle room (Dana et al., 2007; van der Weele et al., 2014). The focus on games with manual choice entry simplifies out-of-sample predictions but precludes experiments with graphical user interfaces (Fisman et al., 2007). Finally, the focus on games with freely available data sets precludes the inclusion of experiments with real-effort tasks preceding the dictator game. However, as reviewed above, the main patterns in real-effort games resemble those in dictator games with generalized endowments and windfall budgets, three of which are included.

A notable difference between the analyzed dictator game experiments concerns the language used in the instructions for assigning the players' endowments. In standard dictator games (e.g. AM02 and HJ06), direct assignments are avoided by stating that "a number of tokens is to be divided", while in taking games (e.g. List07, Bard08, and KMR14), endowments are explicitly assigned prior to the choice task. This may provoke status quo and endowment effects (Samuelson and Zeckhauser, 1988; Kahneman et al., 1991) but to our knowledge it has not been discussed as a behavioral confound in preference analyses of (generalized) dictator games. Table 4 in the appendix reviews the relevant passages in the experimental instructions and distinguishes between neutral language, where specific assignments of the endowments to either of the players are avoided, and loaded language, where initial endowments are specifically assigned or otherwise attributed to either of the players. Neutral language is typically used in standard dictator game experiments (AM02 and HJ06) and in sorting games (LMW12). Loaded language is typically used in experiments with generalized endowments or taking options. The hypothesis that such language differences affect the distribution of reference points and thus induce endowment effects as observed in other studies will be verified below and will be taken into account throughout the entire analysis.

For the following analysis, we use the simplest formulation of reference points that seems conceivable, simplifying even in relation to Assumption 3, in order to rule out any biases in the results due to choosing functional forms,

$$r_1 = w_1 \cdot B_1 + w_2 \cdot tB_2, \qquad r_2 = w_1 \cdot tB_2 + w_2 \cdot B_1.$$

Since the qualitative results hold regardless of the distribution of reference points, such functional form assumptions are largely irrelevant, however. The robustness checks in Appendix C explicitly show that alternative functional forms mapping endowments to reference points yield results very similar to those reported here. As above, we allow that the weights $w_1, w_2 \in [0,1]$ do not necessarily add up to 1, while we assume that they satisfy $w_1 \geq w_2$. The former allows that subjects may be both altruistic givers ($w_1 + w_2 < 1$) and social pressure givers ($w_1 + w_2 \geq 1$), thereby capturing the types observed by DellaVigna et al. (2012), while the latter assumes that subjects put higher weight on the role they end up playing in case their decision turns out to be payoff relevant. As usual, we fix the loss-aversion parameter at $\delta = 2$ to remove a degree of freedom.

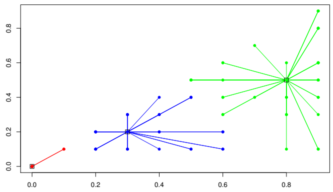
5.2 Heterogeneity and consistency of reference points

First, we examine heterogeneity of reference points within experiments (i.e. within subject pools) and consistency of reference point distributions across experiments (i.e. types of dictator games). We begin with examining consistency across experiments. For, the differences in the language used when assigning endowments potentially preclude consistency across experiments, which might render the subsequent robustness analysis futile. Further, it would limit applicability of reference dependent concepts such as welfare-based altruism, or indeed any existing concept, to understand the behavioral reasons for differences in giving across experiments.

Formally, we estimate the individual reference points of all subjects in the largest experiment from each class of games: dictator games (AM02), games with generalized endowments (KMR13), sorting games (LMW12), and taking games (KMR14). To be precise, we estimate all four individual preference parameters for all subjects, as reference points cannot be estimated without controlling for altruism α and efficiency concerns β, but in the present subsection, we focus on the distributions of reference points. As the estimation procedure is standard maximum likelihood all details on optimization algorithms, generation of starting values, and cross-checking to ensure global optimality of estimates are relegated to the appendix. After estimating the reference point weights (w_1, w_2) for all subjects, we evaluate their structure in a cluster analysis by affinity propagation (Dueck and Frey, 2007). Figure 2 provides the results.

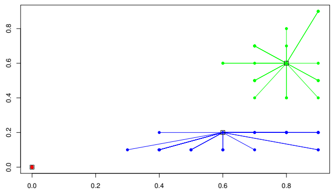
Figure 2: Distribution of reference point weights across types of dictator games

Dictator games (AM02)



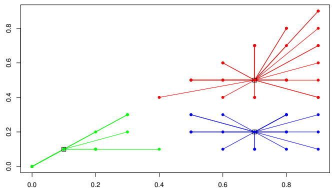
	w_1	w_2	Size	
Cluster 1	0.002	0.002	104/176	59%
Cluster 2	0.359	0.203	32/176	18%
Cluster 3	0.77	0.503	40/176	23%

Generalized endowments (KMR13)



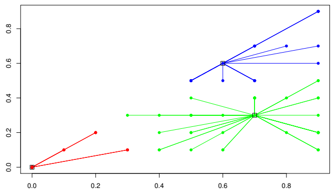
	w_1	w_2	Size	
Cluster 1	0	0	60/119	50%
Cluster 2	0.631	0.144	32/119	27%
Cluster 3	0.774	0.607	27/119	23%

Sorting games (LMW12)



	w_1	w_2	Size	
Cluster 1	0.106	0.084	32/94	34%
Cluster 2	0.71	0.215	20/94	21%
Cluster 3	0.717	0.571	42/94	45%

Taking games (KMR14)



	w_1	w_2	Size	
Cluster 1	0.049	0.032	37/106	35%
Cluster 2	0.69	0.267	39/106	37%
Cluster 3	0.657	0.593	30/106	28%

Note: For the largest experiments from each type of generalized dictator game, all individual reference point weights (w_1, w_2) are estimated, plotted with w_1 on the horizontal axis and w_2 on the vertical axis, and clustered by affinity propagation (Dueck and Frey, 2007). The centers and sizes of the three clusters identified in each case are provided in the respective tables to the right.

Consistently across data sets, three clusters of subjects are identified. All clusters tend to be of similar size, comprising around one third of the subjects. In all cases, there is one group of subjects with endowment-independent reference points ($w_1 \approx w_2 \approx 0$), one group of subjects with "satisfiable" reference points where weights add up to less than one ($w_1 + w_2 < 1$), and one group of subjects with "excessive" reference points where weights add up to more than one ($w_1 + w_2 \geq 1$). The center of the second group moves a little between studies, but overall, the centers and sizes of the clusters are remarkably robust—and they fit received findings in the literature. The first group contains the "egoistic" subjects maximizing their pecuniary payoffs, a group comprising around one third of the subjects in all dictator game experiments. The members of the second and the third group comprise subjects that transfer tokens to the recipients either out of largely altruistic concerns (second group) or out of perceived social pressure (third group)—and further corroborating DellaVigna et al. (2012), these groups are similarly large.[12]

Result 1. *Across all four types of dictator games, there are three similarly-sized groups of subjects: subjects with endowment-independent reference points (mostly egoists), subjects with satisfiable reference points ("altruistic givers"), and subjects with non-satisfiable reference points ("social pressure givers").*

5.3 Significance and robustness of welfare-based altruism

Next, if reference dependence is a *robust* behavioral trait, then accounting for it improves our understanding of giving across contexts, and it does so not only ex-post but also ex-ante. That is, acknowledging reference dependence should improve predictions across contexts, i.e. across types of dictator games. This way, it would help improve policy recommendations and guide (behavioral) mechanism design.[13] Given the data sets analyzed here, we can evaluate this question directly by analyzing predictions across the types of dictator game experiments listed in Table 2.

In addition, if reference dependence is a *behavioral primitive*, then it improves on alternative ways of providing the implied degrees of freedom. Given the existing literature, there are two arguably natural extensions of CES altruism that have to be considered as benchmark models. The first benchmark extends CES altruism by warm glow and cold prickle, as proposed by Korenok et al. (2014). Using (e_1, e_2) as the players' endowments, this model is expressed as

$$u(\pi) = (1-\alpha_1-\alpha_2-\alpha_3)\cdot\pi_1^{\beta}+\alpha_1\,\pi_2^{\beta}+\alpha_2\cdot|e_1-\pi_1|_+^{\beta}-\alpha_3\cdot|e_2-\pi_2|_+^{\beta}. \tag{+ Warm Glow/Cold Prickle}$$

As above, $|x|_+$ equates with x if $x>0$ and it equates with 0 otherwise. Thus, $|e_1-\pi_1|_+$ captures the amount transferred by the dictator from her endowment (inducing "warm glow" which is independent of the amount received by the recipient), and $|e_2-\pi_2|_+$ captures the amount taken from the recipient's endowment (inducing "cold prickle"). The other benchmark extends CES altruism by motives of envy and guilt (Fehr and Schmidt, 1999) as proposed by Korenok et al. (2012).

$$u(\pi) = (1-\alpha_1-\alpha_2-\alpha_3)\cdot\pi_1^{\beta}+\alpha_1\,\pi_2^{\beta}-\alpha_2\cdot|\pi_1-\pi_2|_+-\alpha_3\cdot|\pi_2-\pi_1|_+ \tag{+ Inequity Aversion}$$

[12] Members of both the second and the third group react to the endowments induced via the experimental design. The difference is that the reference points of members in the second group do not eat up the entire budget, while the reference points of members in the third group cannot be satisfied jointly. The members of the third group transfer tokens aiming to satisfy both players' reference points as good as possible, and in this sense, they react solely to the social pressure they perceive due to their (subjective) reference points. The members of the second group, however, react significantly weaker to the social pressure (i.e. to induced endowments), thanks to having smaller weights (w_1, w_2) and mainly decide how to transfer the (often substantial) residual amount after satisfying both reference points. In this sense, they are altruistic givers.

[13] A related range of applications concerns attempts to factor out altruism (or risk aversion) from behavior, to identify other sources of gender-related differences (see for example Niederle and Vesterlund, 2007). To this end, first some treatment is run to identify known preference differences, and these differences are then extrapolated to another task to help identify residual differences. The extrapolation requires predictive accuracy across contexts as analyzed here.

Table 3: Behavioral predictions across types of dictator game experiments

		Descriptive	Predictive	Details on predictions of ...			
Calibrated on	Altruism is ...	Adequacy	Adequacy	Dictator	Endowments	Taking	Sorting
Dictator Games	Payoff based (CES)	1460.9	8950.5	1343.4	4339	2353.3	914.7
	+ Warm Glow/Cold Prickle	1507.3^{--}	8854.6	1343	4218.4^{+}	2375.2	917.9
	+ Inequity Aversion	1234.6^{++}	8794.8^{++}	1217.1^{+}	4311.7	2360.7	905.3
	Welfare based	1146.6^{++}	8758^{++}	1279.8^{+}	4273.8^{+}	2316.6^{+}	887.7
	Welfare based (adj)	1146.6^{++}	8603.9^{++}	1263.9^{+}	4152.5^{++}	2300.8^{++}	888.2
Gen Endowments	Payoff based (CES)	2896.6	8752.9	4260.4	826.1	2613.8	1052.7
	+ Warm Glow/Cold Prickle	2395.5^{++}	8967.8^{--}	4289.6	954.5^{--}	2649.7	1074
	+ Inequity Aversion	2800.1^{+}	8916.4^{--}	4333.6^{-}	849.9	2663^{--}	1069.9^{--}
	Welfare based	2662.7^{++}	8416.7^{++}	4084.2^{++}	767.9^{+}	2565.9^{+}	998.7^{++}
	Welfare based (adj)	2662.7^{++}	7867.7^{++}	3985.8^{++}	637.1^{++}	2351^{++}	895.4^{++}
Taking Games	Payoff-based (CES)	1482.4	9700.7	3739.3	4466.7	579.7	914.9
	+ Warm Glow/Cold Prickle	1451.8	10252.5^{--}	4263.8^{--}	4408.7	592.8	987.2^{--}
	+ Inequity Aversion	1419.2^{+}	9736.7	3543.3^{++}	4698.2^{--}	576.6	918.5
	Welfare-based	1226.4^{++}	9499.7^{+}	3729.2	4343.2	568.5^{+}	858.8^{++}
	Welfare based (adj)	1226.4^{++}	9270.3^{++}	3633^{+}	4232.9^{++}	559.3^{++}	846.6^{++}
Aggregate	Payoff based (CES)	5839.8	27404.1	9343.1	9631.8	5546.8	2882.4
	+ Warm Glow/Cold Prickle	5354.6^{++}	28075^{--}	9896.5^{--}	9581.6	5617.8^{-}	2979.1^{--}
	+ Inequity Aversion	5453.9^{++}	27447.9	9094^{+}	9859.8^{--}	5600.4^{--}	2893.7
	Welfare based	5035.7^{++}	26674.4^{++}	9093.2^{++}	9385^{++}	5451^{++}	2745.2^{++}
	Welfare based (adj)	5035.7^{++}	25740.4^{++}	8883.6^{++}	9023.5^{++}	5212.2^{++}	2631.2^{++}

Note: For each type of dictator game experiment used to estimate the parameters (standard "Dictator games" in AM02, "Generalized endowments" in KMR13, "Taking Games" in KMR14), we report for each of the five models the in-sample fit ("Descriptive Adequacy"), the pooled out-of-sample fit by predicting all other experiments in Table 2 ("Predictive Adequacy"), and the detailed predictive adequacy for each type of experiments as distinguished in Table 2 (the four right-most columns). Plus and Minus signs indicate significance of differences of the Akaike Information Criterion (AIC) for each of the generalizations of the CES model to the CES model. The likelihood-ratio tests (Schennach and Wilhelm, 2016) are robust to misspecification and arbitrary nesting, and we distinguish significance levels of .05 ($^{+}$,$^{-}$) and .01 ($^{++}$,$^{--}$). In all cases, we cluster at the subject level to account for the panel character of the data.

An attractive feature of these models is that they also contain four free parameters in total, in this respect equating with welfare-based altruism, which implies that these models can be estimated following the exact same procedure as welfare-based altruism. This way, we can ensure comparability of the results. All the technical details on likelihood maximization and statistical tests are provided in the appendix.

We estimate all models on each of the three largest data sets, i.e. on standard dictator games (AM02), on games with generalized endowments (KMR13), and on games with taking options (KMR14), and predict behavior in all data sets listed in Table 2.[14] The results are summarized in Table 3. The "Descriptive Adequacy" is the Akaike information criterion of the in-sample fit, i.e. the sum of absolute value of the log-likelihood and number of parameters (in-sample, every reference point of every subject counts as a free parameter). The predictive adequacy is reported both in aggregate (column "Predictive Adequacy") and segregated by type of dictator game to be

[14] We do not consider predictions based on estimates from the sorting game experiment of LMW12, as their experimental design varies neither the transfer rate (fixed to 1 : 1) nor the endowments of dictators and receivers, varying only the price for sorting out. This way, the preference parameter β, capturing the preference for efficiency and equity, is not identified and predictions are largely uninformative.

predicted (sets of columns "Details on predictions of . . . "). In all cases, descriptive and predictive adequacies are reported for each of the four models discussed so far, payoff-based CES altruism, the extensions additionally allowing for either warm glow and cold prickle or envy and guilt, and the welfare-based altruism model. In addition, we report results from a robustness check allowing for variations in the strength of assignments of endowments, the model "Welfare based (adj)" that we discuss below. Finally, in the lower part of Table 3, all the numbers in the upper part are aggregated across all three in-sample data sets to provide the overall picture.

Descriptive adequacy First, we examine the in-sample fit (column "Descriptive Adequacy"). In aggregate, all generalized models significantly improve on the payoff-based CES model despite accounting for the additional parameters using AIC. The proposed model of welfare-based altruism is unique in that it improves highly significantly upon CES in all three contexts and in this sense represents the only robustly fitting model. Yet, the observation that on aggregate all three models do so suggests that perhaps they all capture differently important but significant facets of behavior. If so, this will show in their predictive adequacy.

Predictive adequacy Evaluating robustness of the explanatory power (column "Predictive Adequacy") changes the picture substantially. Welfare-based altruism improves on CES' predictions in all contexts, regardless of the data set used for estimation, and mostly significantly so. That is, regardless of the context the model is fitted on and of the class of dictator game experiments to be predicted, the resulting goodness-of-fit is higher than that of the standard CES model, in all 3×4 cases, significantly so in $9/12$ cases, and always on aggregate.[15] The explanatory power of reference dependence in giving may therefore be considered robust.

At the other extreme, extending CES altruism by warm glow and cold prickle predicts behavior better than CES in only $3/12$ cases but worse than CES in $9/12$ cases. On aggregate, its predictions are significantly worse than CES, and this obtains although warm glow and cold prickle seem to capture behavior (in-sample) in the case of generalized endowments best. This applies only in-sample, however, even predictions for the other experiments allowing for generalized endowments fit worse than CES (and all other models), suggesting that the extension allowing for warm glow and cold prickle does not capture a robust behavioral trait in the games analyzed here.

Finally, the extension allowing for envy and guilt ("inequity aversion") is in-between with respect to its descriptive and predictive adequacy. While it fits worse than welfare-based altruism in all contexts, both in-sample and out-of-sample, at least it does not overfit on aggregate and thereby it improves on warm glow and cold prickle. That is, on aggregate, accounting for envy and guilt does not yield predictions that are significantly worse than not doing so (as in the standard CES model). Nonetheless, predictions also do not improve on aggregate, suggesting that envy and guilt are actually not robust behavioral traits in giving—they allow to rationalize Leontief choices, but those are not robustly chosen.[16] Corroborating this observation, if we evaluate predictions

[15] Note that, as mentioned in the notes to all tables and in the appendix, we use the Schennach-Wilhelm likelihood ratio test throughout (Schennach and Wilhelm, 2016), clustered at the subject level. It is robust to misspecification of models, arbitrary nesting structures, and captures the panel character of the data with multiple observations per subject.

[16] In particular in the games with generalized (non-zero) endowments, the payoff-equalizing "Leontief" option happens to be rarely chosen (Korenok et al., 2013). For example, only $2/116$ subjects in KMR14 are strict Leontief types, whereas around 20% of the subjects are in standard dictator games (see AM02). In this context, predictions assuming that envy and guilt are behavioral factors fit poorly.

across all 4×3 cases, inequity aversion's predictions significantly improve on CES in $2/12$ cases, it predicts significantly worse in $4/12$ cases, and overall, its predictive adequacy is slightly worse than that of the payoff-based CES model.

Result 2. *Welfare-based altruism improves on CES altruism for all types of DG experiments, both descriptively (in-sample) and robustly (out-of-sample) highly significantly. None of the benchmark models does so in more than* $2/12$ *cases, corroborating the theoretical prediction that reference dependence is a causal factor in giving across contexts.*

Table 3 additionally informs on a robustness check accounting for the variation in language used assigning endowments (Table 4 in the appendix). In this robustness check, we allow for homogeneous shifts in weights between experiments, by introducing a free parameter per set of predictions. Assuming the in-sample estimates of the weights are (w_1, w_2), we allow the out-of-sample weights to be (w_1^γ, w_2^γ), where the shift $\gamma \geq 0$ is homogeneous for all subjects. With $\gamma < 1$ all weights increase and with $\gamma > 1$ all weights decrease—reflecting stronger and weaker assignments, respectively. Introducing γ as a free parameter allows us to either strengthen or weaken weights homogeneously for all subjects. Naturally, this has no effect in-sample, but it has substantial effects out-of-sample—amounting to around 1000 points on the log-likelihood scale in total (yielding a drop from 26674.4 to 25740.4). This improvement is highly significant given the low number of additional parameters used, strongly underlining the initial hypothesis that the language used in experimental instructions is highly relevant in shaping behavior. The present analysis is not suited nor intended to fully clarify the relevance of language used assigning endowments, but changes in language across experiments, which have not been explicitly discussed in the literature on generalized dictator games, are evidently not innocent choices in experimental design. This does not directly affect the above results, since acknowledging language differences as a factor shaping reference points only strengthens the case for welfare-based altruism, but such differences may be acknowledged more explicitly when designing and analyzing future experiments.

5.4 Relation to social norms and "social appropriateness"

Starting with Krupka and Weber (2013), a growing literature relates giving observed in experiments to norm compliance. Subjects are assumed to have a common understanding of the "social appropriateness" of options, which in turn affects dictator behavior and is a function of the social norms applying in a given context. In a novel experimental design, Krupka and Weber measure social appropriateness by having (third) subjects play a coordination game—asking each subject how "socially appropriate" the available options are in the eyes of their co-players and paying a prize to all subjects picking the modal response. The mean of all appropriateness ratings is mapped into a measure $s_x \in [-1, 1]$ for all options x, with $s_x = -1$ indicating highly inappropriate and $s_x = 1$ indicating highly appropriate options. Krupka and Weber then examine if a utility function of the form

$$u_x = \pi_x + \alpha s_x \tag{4}$$

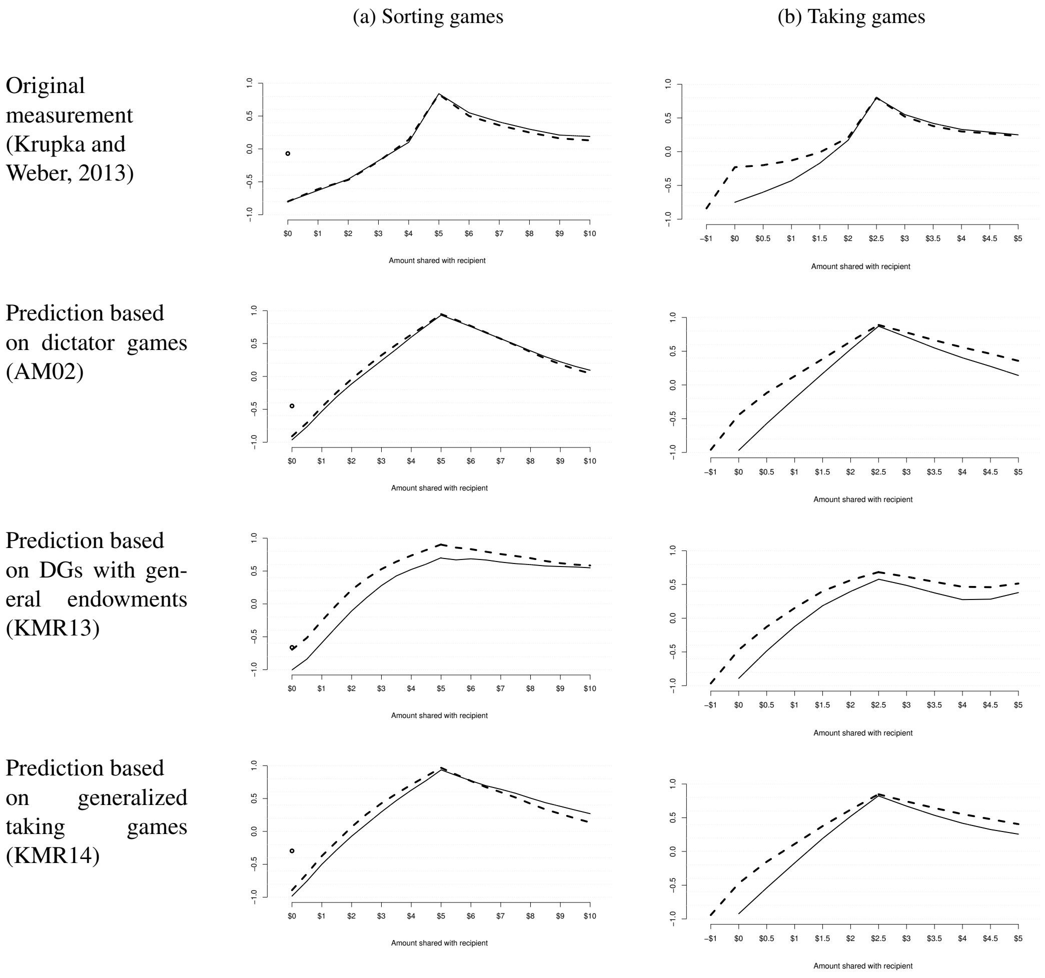
fits behavior observed in earlier dictator game experiments, using the weight α as a free parameter. While statistical tests supporting the results are not provided, the plots in Krupka and Weber (2013) suggest a good fit after calibrating α. This finding has been interpreted as indicating that behavior is norm-guided, rather than being payoff or welfare concerned as assumed in earlier work. In the following, we clarify the relation of our findings to those of Krupka and Weber (2013) and subsequent work, to discuss how we may think of welfare-based altruism as a foundation of norm-guided giving.

To this end, let us recap two main results. Krupka and Weber convincingly demonstrate that experimental subjects are able to predict behavior in taking and sorting games, a feat that existing behavioral models struggled to achieve. We have shown that welfare-based altruism also allows to predict behavior, and hence our conjecture: the two are likely to correlate. A post-hoc straightforward approach would be to take our predictions of utility u_x across options, the respectively induced payoffs π_x, and to then compute social appropriateness s_x by inverting Eq. (4) for all options x. We skip this fairly unintelligible exercise and evaluate whether social appropriateness may be deduced from first principles.

Krupka and Weber (2013) interpret social appropriateness as reflecting the social norm that dictators facing a specific dictator game trade off with their self-interest. They argue that since their elicitation method (i) makes uninvolved subjects rate actions rather than outcomes and (ii) incentivizes subjects to rate in accordance with what they regard as a socially shared assessment, the resulting appropriateness ratings satisfy the two main characteristics of a social norm as defined by Elster (1989). These defining features of social norms are closely related to the "social contract" of Rawls (1971), which specifies a standard for social and distributive justice that "free and rational persons concerned to further their own interests would accept in an initial position of equality" (p. 11). The idea is that the members of a society would unanimously agree to the social contract if they met behind the "veil of ignorance", a hypothetical place where they are unaware of their positions in society (see also Konow, 2003). According to Rawls (1971) the social contract emerging in such a situation would prescribe a distribution that equalizes individual welfares unless inequality is to the advantage of the individual with the minimum welfare. While for obvious reasons an experimental test of Rawls hypothesis can never be perfect, Krupka and Weber's subjects share some central characteristics with Rawls' society members behind the veil of ignorance. They can be thought of as impartial since they are uninvolved while they are part of the same society as the involved players. Furthermore, they are incentivized to find an agreement instead of simply voicing their opinions. Therefore, looking at Krupka and Weber's social appropriateness ratings through the lense of our welfare-based altruism model allows us to test the Rawlsian hypothesis of social welfare being the minimum of all individual welfares, joint with the assertion that social appropriateness simply transforms social welfare to a scale ranging from highly inappropriate (-1) to highly appropriate (1).

Since our welfare-based approach directly builds on individual welfares v_1 and v_2, we are able to directly test the asserted Rawlsian link between appropriateness and welfares—simply by predicting individual welfares for all options in the sorting and taking games analyzed by Krupka and Weber, taking the minimum of v_1 and v_2 across options, and rescaling such that a measure ranging from -1 to $+1$ results. Specifically, we predict the social appropriateness ratings for both taking and sorting games analyzed by Krupka and Weber based on our estimates from each of

Figure 3: Relation of experimentally measured “social appropriateness” (Krupka and Weber) to the Rawlsian prediction following from our estimates


(a) Sorting games

(b) Taking games

Original measurement (Krupka and Weber, 2013)

Prediction based on dictator games (AM02)

Prediction based on DGs with general endowments (KMR13)

Prediction based on generalized taking games (KMR14)

(c) Correlation between observed and predicted appropriateness

	Sorting games		Taking games	
Predictions based on …	Spearman-ρ	p-value	Spearman-ρ	p-value
Dictator games (AM02)	0.641	(0.001)	0.738	(0)
Gen endowments (KMR13)	0.667	(0.001)	0.766	(0)
Taking games (KMR14)	0.644	(0.001)	0.751	(0)

Note: The “sorting games” compare appropriateness in a standard dictator game with endowments of 10 for the dictator and 0 for the recipient to appropriateness in a sorting game where the dictator game is succeeded by giving the dictator the option to sort out at costs of 1. The “taking games” compare appropriateness in a standard dictator game with endowments of 10 for the dictator and 5 for the recipient to appropriateness in a taking game where the dictator game may alternatively take one currency unit from the recipient’s endowment. The plots follow Krupka and Weber: solid lines represent the social appropriateness in the standard dictator games and dashed lines represent social appropriateness in the sorting and taking games, respectively. The single “dot” in the sorting games reflects the appropriateness of sorting out.

the three experiments analyzed before (AM02, MKR13, and KMR14). This yields 3×2 profiles of appropriateness ratings, which we then relate to the measurements of Krupka and Weber.[17] The results are reported in Figure 3 and strongly corroborate the relation of social appropriateness and Rawlsian welfare asserted already by Krupka and Weber. The correlation between the out-of-sample predictions and the in-sample measurements of Krupka and Weber is very high, around 0.65 in sorting games and around 0.75 in taking games, regardless of the data set which the prediction is based on. We therefore conclude as follows.

Result 3. *Krupka and Weber's measure of social appropriateness strongly correlates with the Rawlsian notion of welfare, based on out-of-sample predictions of individual welfares derived from the above model of welfare-based altruism.*

That is, social appropriateness is founded in welfare concerns in the intuitive Rawlsian manner alluded to by Krupka and Weber. It seems futile to ask which came first, welfare concerns or social appropriateness/social norms, they rather appear to be two sides of the same coin. The received interpretation that giving reflects context-dependent social norms rather than more fundamental payoff and welfare concerns seems premature, but so would the opposite. From a practical point of view, both approaches seem to have distinctive strengths. Analyses relating behavior to social appropriateness need not be concerned with individual preferences and can focus on the picture at large. In turn, the behavioral foundation in welfare concerns has an independent axiomatic foundation in established behavioral principles, which greatly facilitates application across contexts, and the implied S-shape of individual welfares has been observed in many contexts, which promises reliable predictions and policy recommendations out-of-sample.

6 Conclusion

This paper contributes to the efforts in reorganizing models of the interdependence of preferences (List, 2009; Malmendier et al., 2014) that was initiated by a wave of generalized dictator game experiments allowing for non-trivial endowments (Bolton and Katok, 1998; Korenok et al., 2013), taking options (List, 2007; Bardsley, 2008), and sorting options (Dana et al., 2006; Lazear et al., 2012). The new observations were interpreted as being incompatible with observations from standard dictator games and in the existing literature a plethora of approaches have been proposed to capture them: menu dependent preferences and cold prickle to capture taking decisions, warm glow and social norms to capture endowment effects, image concerns and social pressure to capture sorting decisions. Considering this range of proposals simply to organize observations on giving under complete information, robustly applicable models of this most fundamental of economic activities appear to be out of reach (Korenok et al., 2014)—illustrating a surprisingly tight bound on economic modeling. However, following our basic intuition, we suspected that this apparent

[17] Specifically, for each subject in our in-sample experiments (AM02, KMR13, KMR14), we determine the individual welfares if that subject would play either role, v_1 and v_2. We then assume that an impartial observer in the sense of Krupka and Weber determines appropriateness as follows: Across dictators, what is their average individual welfare from choosing x conditional on choosing x in the first place. Across recipients, what is their average individual welfare from getting x conditionally on being confronted with x in the first place (which is an empty condition, stated only for symmetry). The lesser of these conditional expectations is the unscaled Rawlsian appropriateness of each option, and rescaling to $[-1,1]$ across options yields our out-of-sample prediction for Krupka-Weber appropriateness.

incompatibility can be resolved once we acknowledge non-convexity and reference dependence of preferences—as known from choice under risk.

We propose a model that entails exactly this, following an analysis that differs from earlier work in four important ways. First, we start with an axiomatic foundation based on which we characterize a general family of utility representations capturing interdependence of preferences. This identifies the class of candidate models. Second, we complement the theoretical analysis by a comprehensive econometric analysis of model validity across nine laboratory experiments to provide a rigorous, objective assessment of model adequacy. Third, as a technical innovation in the axiomatic derivation, we formally distinguish contexts, which allows us to formalize the notion of narrow bracketing as a property of preferences, and thus to establish a formally tight but ex-ante unsuspected link between four large literatures in behavioral economics: prospect theory (Kahneman and Tversky, 1979), narrow bracketing (Read et al., 1999), altruism (Andreoni and Miller, 2002), and reference dependence (Kőszegi and Rabin, 2006). Finally, our results reconcile a wide range of seemingly inconsistent experimental results with approaches and results from classical decision theory.

Implicitly, instead of constructing a utility functional that fits as many stylized facts as possible, we derive a utility representation from established behavioral principles such as scaling invariance and narrow bracketing. The theoretical predictions about behavior in generalized dictator games, the tight relations to four major branches of behavioral economics, and the fact that welfare-based altruism directly formalizes the widespread notion that altruism is a concern for the welfare of others, while being derived from universal behavioral axioms not related to altruism or dictator games, renders this a promising model for future work. Our econometric results on in-sample and out-of-sample adequacy provide validity in this respect, and both the model's generality and its quantitative adequacy open up a number of exciting avenues for future research.

These include experimental analyses of preferences and reference points, based on an axiomatically solid and econometrically validated model, theoretical analyses of utility representations under alternative axioms and of revealed preference with non-convexities (see also Halevy et al., 2017), empirical and theoretical analyses of behavioral welfare and preference laundering,[18] and, exploiting the relation to choice under risk, behavioral analyses of giving under incomplete information (as in Dana et al., 2007, and Andreoni and Bernheim, 2009) or in multilateral interactions. Due to the large extent of similarity of charitable giving and dictator behavior in the laboratory (Konow, 2010; Huck and Rasul, 2011; DellaVigna et al., 2012), a particularly immediate range of applications lies in structural analyses of charitable giving (DellaVigna, 2009; Card et al., 2011) generalizing, for example, the work of DellaVigna et al. (2012, 2016) and Huck et al. (2015).

References

Aczél, J. (1966). *Lectures on functional equations and their applications*, volume 19. Academic press.

[18] Letting all agents have equal weight, our analysis establishes a utilitarian welfare function which contains Rawls and Harsanyi as special cases (for $\beta \to -\infty$ and $\beta = 1$, respectively), where individual welfares are the prospect-theoretic utilities from single-person decision making. This provides an axiomatic foundation for preference laundering in welfare analyses (Goodin, 1986), i.e. to disregard concerns for others (such as envy) in behavioral welfare economics, which drastically affects policy recommendations (see also Piacquadio, 2017).

Almås, I., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2010). Fairness and the development of inequality acceptance. *Science*, 328(5982):1176–1178.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence. *Journal of Political Economy*, pages 1447–1458.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal*, 100(401):464–477.

Andreoni, J. (1995). Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics*, 110(1):1–21.

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

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

Andreoni, J., Rao, J. M., and Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy*, 125(3):625–653.

Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1):544–555.

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

Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, 11(2):122–133.

Becker, G. S. (1974). A theory of social interactions. *Journal of political economy*, 82(6):1063–1093.

Bellemare, C., Kröger, S., and van Soest, A. (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, 76(4):815–839.

Bellemare, C., Sebald, A., and Strobel, M. (2011). Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models. *Journal of Applied Econometrics*, 26(3):437–453.

Bolton, G. E. and Katok, E. (1998). An experimental test of the crowding out hypothesis: The nature of beneficent behavior. *Journal of Economic Behavior & Organization*, 37(3):315–331.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, pages 166–193.

Breitmoser, Y. (2013). Estimation of social preferences in generalized dictator games. *Economics Letters*, 121(2):192–197.

Breitmoser, Y. (2017). Discrete choice with representation effects. CRC TRR 190 Working Paper.

Broberg, T., Ellingsen, T., and Johannesson, M. (2007). Is generosity involuntary? *Economics Letters*, 94(1):32–37.

Brock, J. M., Lange, A., and Ozbay, E. Y. (2013). Dictating the risk: Experimental evidence on giving in risky environments. *American Economic Review*, 103(1):415–437.

Camerer, C., Cohen, J., Fehr, E., Glimcher, P., and Laibson, D. (2017). *Neuroeconomics*, volume 2, pages 153–217. Princeton University Press, editors: john kagel and alvin roth edition.

Camerer, C. and Ho, T.-H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874.

Camerer, C. F., Ho, T.-H., and Chong, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898.

Cappelen, A. W., Halvorsen, T., Sørensen, E. Ø., and Tungodden, B. (2017). Face-saving or fair-minded: What motivates moral behavior? *Journal of the European Economic Association*, 15(3):540–557.

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

Cappelen, A. W., Moene, K. O., Sørensen, E. Ø., and Tungodden, B. (2013a). Needs versus entitlements an international fairness experiment. *Journal of the European Economic Association*, 11(3):574–598.

Cappelen, A. W., Nielsen, U. H., Sørensen, E. Ø., Tungodden, B., and Tyran, J.-R. (2013b). Give and take in dictator games. *Economics Letters*, 118(2):280 – 283.

Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2010). Responsibility for what? fairness and individual responsibility. *European Economic Review*, 54(3):429–441.

Card, D., DellaVigna, S., and Malmendier, U. (2011). The role of theory in field experiments. *The Journal of Economic Perspectives*, 25(3):39–62.

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

Cherry, T. L. (2001). Mental accounting and other-regarding behavior: Evidence from the lab. *Journal of Economic Psychology*, 22(5):605–615.

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

Cherry, T. L. and Shogren, J. F. (2008). Self-interest, sympathy and the origin of endowments. *Economics Letters*, 101(1):69–72.

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

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

Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and human decision Processes*, 100(2):193–201.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

De Bruyn, A. and Bolton, G. E. (2008). Estimating the influence of fairness on bargaining behavior. *Management Science*, 54(10):1774–1791.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–372.

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *Quarterly Journal of Economics*, 127:1–56.

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2016). Estimating social preferences and gift exchange at work. Technical report, National Bureau of Economic Research.

Dueck, D. and Frey, B. J. (2007). Non-metric affinity propagation for unsupervised image categorization. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and economic behavior*, 47(2):268–298.

Easterlin, R. A. (2001). Income and happiness: Towards a unified theory. *Economic Journal*, 111(473):465–484.

Eckel, C. C., Grossman, P. J., and Johnston, R. M. (2005). An experimental test of the crowding out hypothesis. *Journal of Public Economics*, 89(8):1543–1560.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives*, 3(4):99–117.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. B., and Sunde, U. (2017). Global evidence on economic preferences. Technical report, National Bureau of Economic Research.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and economic behavior*, 54(2):293–315.

Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *The American Economic Review*, 101(2):470–492.

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

Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, 97(5):1858–1876.

Fleurbaey, M. and Maniquet, F. (2011). *A theory of fairness and social welfare*, volume 48. Cambridge University Press.

Gächter, S., Herrmann, B., and Thöni, C. (2004). Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. *Journal of Economic Behavior and Organization*, 55(4):505–531.

Gill, D. and Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *The American economic review*, 102(1):469–503.

Goodin, R. E. (1986). Laundering preferences. In *Foundations of social choice theory*. Cambridge University Press Cambridge.

Gul, F. and Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6):1403–1435.

Halevy, Y., Persitz, D., Zrill, L., et al. (2017). Parametric recoverability of preferences. *Journal of Political Economy*.

Harless, D. W., Camerer, C. F., et al. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62(6):1251–1289.

Harrison, G. W. and Johnson, L. T. (2006). Identifying altruism in the laboratory. In *Experiments Investigating Fundraising and Charitable Contributors*, pages 177–223. Emerald Group Publishing Limited.

Hey, J. D., Lotito, G., and Maffioletti, A. (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of risk and uncertainty*, 41(2):81–111.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7(3):346–380.

Hoffman, E., McCabe, K., and Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *American Economic Review*, 86(3):653–660.

Huck, S. and Rasul, I. (2011). Matched fundraising: Evidence from a natural field experiment. *Journal of Public Economics*, 95(5):351–362.

Huck, S., Rasul, I., and Shephard, A. (2015). Comparing charitable fundraising schemes: Evidence from a natural field experiment and a structural model. *American Economic Journal: Economic Policy*, 7(2):326–69.

Jakiela, P. (2011). Social preferences and fairness norms as informal institutions: experimental evidence. *American Economic Review*, 101(3):509–513.

Jakiela, P. (2015). How fair shares compare: Experimental evidence from two cultures. *Journal of Economic Behavior & Organization*, 118:40–54.

Johnson, N. D. and Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, 32(5):865–889.

Kahneman, D., Knetsch, J. L., and Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, pages 728–741.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1):193–206.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, 90(4):1072–1091.

Konow, J. (2003). Which is the fairest one of all? a positive analysis of justice theories. *Journal of Economic Literature*, 41(4):1188–1239.

Konow, J. (2010). Mixed feelings: Theories of and evidence on giving. *Journal of Public Economics*, 94(3):279–297.

Korenok, O., Millner, E. L., and Razzolini, L. (2012). Are dictators averse to inequality? *Journal of Economic Behavior & Organization*, 82(2):543–547.

Korenok, O., Millner, E. L., and Razzolini, L. (2013). Impure altruism in dictators' giving. *Journal of Public Economics*, 97:1–8.

Korenok, O., Millner, E. L., and Razzolini, L. (2014). Taking, giving, and impure altruism in dictator games. *Experimental Economics*, 17(3):488–500.

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

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

Kőszegi, B. and Rabin, M. (2009). Reference-dependent consumption plans. *American Economic Review*, 99(3):909–936.

Krawczyk, M. and Le Lec, F. (2010). Give me a chance! an experiment in social decision under risk. *Experimental Economics*, 13(4):500–511.

Kritikos, A. and Bolle, F. (2005). Utility-based altruism: evidence from experiments. In *Psychology, Rationality and Economic Behaviour*, pages 181–194. Springer.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Lazear, E. P., Malmendier, U., and Weber, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, 4(1):136–163.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, 1(3):593–622.

List, J. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3):482–493.

List, J. (2009). Social preferences: Some thoughts from the field. *Annual Review of Economics*, 1(1):563–583.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. John Wiley and sons.

Malmendier, U., te Velde, V. L., Weber, R. A., et al. (2014). Rethinking reciprocity. *Annual Review of Economics*, 6(1):849–874.

McCullough, B. D. and Vinod, H. D. (2003). Verifying the solution from a nonlinear solver: A case study. *American Economic Review*, 93(3):873–892.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *Quarterly Journal of Economics*, 122(3):1067–1101.

Oosterbeek, H., Sloof, R., and Van De Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, 7(2):171–188.

Oxoby, R. J. and Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, 65(3):703–713.

Padoa-Schioppa, C. (2009). Range-adapting representation of economic value in the orbitofrontal cortex. *The Journal of Neuroscience*, 29(44):14004–14014.

Padoa-Schioppa, C. and Rustichini, A. (2014). Rational attention and adaptive coding: a puzzle and a solution. *American Economic Review*, 104(5):507–513.

Piacquadio, P. G. (2017). A fairness justification of utilitarianism. *Econometrica*, 85(4):1261–1276.

Powell, M. (2006). The newuoa software for unconstrained optimization without derivatives. *Large-Scale Nonlinear Optimization*, pages 255–297.

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

Rabin, M. and Weizsäcker, G. (2009). Narrow bracketing and dominated choices. *American Economic Review*, 99(4):1508–1543.

Rawls, J. (1971). *A theory of justice*. Harvard university press.

Read, D., Loewenstein, G., and Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, 19(1-3):171–197.

Rohde, K. I. (2010). A preference foundation for fehr and schmidt's model of inequity aversion. *Social Choice and Welfare*, 34(4):537–547.

Rubinstein, A. (2012). *Lecture notes in microeconomic theory: the economic agent*. Princeton University Press.

Ruffle, B. J. (1998). More is better, but fair is fair: Tipping in dictator and ultimatum games. *Games and Economic Behavior*, 23(2):247–265.

Saito, K. (2013). Social preferences under risk: equality of opportunity versus equality of outcome. *American Economic Review*, 103(7):3084–3101.

Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1):7–59.

Schennach, S. and Wilhelm, D. (2016). A simple parametric model selection test. *Journal of the American Statistical Association*.

Schmidt, U. (2003). Reference dependence in cumulative prospect theory. *Journal of Mathematical Psychology*, 47(2):122–131.

Simonsohn, U. and Gino, F. (2013). Daily horizons: evidence of narrow bracketing in judgment from 10 years of mba admissions interviews. *Psychological science*, 24(2):219–224.

Skiadas, C. (2013). Scale-invariant uncertainty-averse preferences and source-dependent constant relative risk aversion. *Theoretical Economics*, 8(1):59–93.

Skiadas, C. (2016). Scale or translation invariant additive preferences. Unpublished manuscript.

Tremblay, L. and Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, 398(6729):704–708.

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, pages 1039–1061.

van der Weele, J. J., Kulisa, J., Kosfeld, M., and Friebel, G. (2014). Resisting moral wiggle room: how robust is reciprocal behavior? *American Economic Journal: Microeconomics*, 6(3):256–264.

Wakker, P. and Tversky, A. (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty*, 7(2):147–175.

Wakker, P. P. (1989). *Additive representations of preferences: A new foundation of decision analysis*, volume 4. Springer Science & Business Media.

Wakker, P. P. and Zank, H. (2002). A simple preference foundation of cumulative prospect theory with power utility. *European Economic Review*, 46(7):1253–1271.

Wilcox, N. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In Cox, J. C. and Harrison, G. W., editors, *Risk aversion in experiments*, volume 12 of *Research in experimental economics*, pages 197–292. Emerald Group Publishing Limited.

Wilcox, N. T. (2011). Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, 162(1):89–104.

Wilcox, N. T. (2015). Error and generalization in discrete choice under risk. *Working paper*.

Appendix

Welfare-based altruism

Yves Breitmoser and Pauline Vorjohann
Humboldt University Berlin

A Relegated proofs

A.1 Proof of Proposition 1

Step 1 Existence of a continuous, additively separable utility representation.
Axioms 1–2 imply existence of a continuous utility representation (see e.g. Rubinstein, 2012, chap. 4). In addition Axiom 3 implies existence of an additively separable utility representation, see Theorem III.4.1 in Wakker (1989) for each context $\pi \in \Pi$. That is, there exists a family of functions $\{v_{\pi,i} : \mathbb{R} \to \mathbb{R}\}_{\pi\in\Pi, i\le n}$ such that $\pi(x) \succsim_\pi \pi(y) \Leftrightarrow u_\pi(x) \ge u_\pi(y)$ for all $x,y \in X$ and $\pi \in \Pi$ with

$$u_\pi(x') = \sum_{i\le n} v_{\pi,i}\big(\pi_i(x')\big) \tag{5}$$

for all $x' \in X, \pi \in \Pi$. For later reference, Wakker's Theorem III.4.1 also establishes that all additively separable representations $\tilde{u}_\pi$ of $\succsim_\pi$ are positive affine transformations of one another. Also note that the representations obtained so far may be context dependent.

Step 2 Context independent (v_i) by narrow or broad bracketing.
We show that additionally assuming either Axiom 5 or Axiom 6 implies that there exists a family of functions $\{v_i : \mathbb{R} \to \mathbb{R}\}_{i\le n}$ and $r : \Pi \to \mathbb{R}^n$ such that

$$u_\pi(x) = \sum_{i\le n} v_i\big(\pi_i(x)\big) \tag{Broad bracketing}$$

$$u_\pi(x) = \sum_{i\le n} v_i\big(\pi_i(x) - r_i(\pi)\big) \tag{Narrow bracketing}$$

represent $\succsim_\pi$ for all $\pi \in \Pi$.

Narrow bracketing: Fix any $\pi \in \Pi$, any $x \in X$. We show that if the preferences obey Axioms 1-3 and Axiom 6, then they admit the claimed representation for any function $r : \Pi \to \mathbb{R}^n$ with $r(\pi') = \pi'(x) - \pi(x)$ for all $\pi' \in \Pi$. Fix this r and any $\pi' \in \Pi$. By Assumption 1.1, $r(\pi') = \pi'(x') - \pi(x')$ for all $x' \in X$. Also note $r(\pi) = \mathbf{0}$. By narrow bracketing, we know that $\succsim_\pi$ is equivalent to $\succsim_{\pi'}$, and using the utility functions obtained in Step 1, this implies

$$\sum_{i\le n} v_{\pi,i}\big(\pi_i(x)\big) \ge \sum_{i\le n} v_{\pi,i}\big(\pi_i(y)\big) \quad \Leftrightarrow \quad \sum_{i\le n} v_{\pi',i}\big(\pi'_i(x)\big) \ge \sum_{i\le n} v_{\pi',i}\big(\pi'_i(y)\big) \tag{6}$$

for all $x,y \in X$. Since $\pi(x) = \pi'(x) - r_i(\pi')$ for all x by construction of r, this yields

$$\sum_{i\le n} v_{\pi,i}\big(\pi'_i(x) - r_i(\pi')\big) \ge \sum_{i\le n} v_{\pi,i}\big(\pi'_i(y) - r_i(\pi')\big) \quad \Leftrightarrow \quad \sum_{i\le n} v_{\pi',i}\big(\pi'_i(x)\big) \ge \sum_{i\le n} v_{\pi',i}\big(\pi'_i(y)\big)$$

for all $x,y \in X$. Since this holds true for all $\pi' \in \Pi$, the claim is established using $v_i = v_{\pi,i}$ for all $i \leq n$. Note again that u (and thus v) is unique up to positive affine transformation, and that for any π', if $\pi' = \pi + c$, then $r(\pi') = r(\pi + c) = r(\pi) + c$ by construction.

Broad bracketing: For each context π, fix value functions $(v_{\pi,i})$ representing $\succsim_\pi$ as obtained in Step 1. By broad bracketing, value functions $(\tilde{v}_\pi)_{\pi \in \Pi}$ representing preferences exist such that for all $x, x' \in X$ and all $\pi, \pi' \in \Pi$,

$$\pi(x) = \pi'(x') \qquad \Leftrightarrow \qquad \sum_{i \leq n} \tilde{v}_{\pi,i}\big(\pi_i(x)\big) = \sum_{i \leq n} \tilde{v}_{\pi',i}\big(\pi'_i(x')\big).$$

Given any such family $(\tilde{v}_{\pi,i})$, and using $P = \cup_{\pi \in \Pi} \pi[X]$, define the functions $\{v_i : P \to \mathbb{R}\}_{i \leq n}$ such that for all $p \in P$,

$$v_i(p_i) = \tilde{v}_{\pi,i}(\pi_i(x)) \qquad \text{for some } (\pi, x) : p = \pi(x).$$

Adequate (π, x) exist for all $p \in P$ by construction of P. By broad bracketing, this implies

$$v_i(p_i) = \tilde{v}_{\pi,i}(\pi_i(x)) \qquad \text{for all } (\pi, x) : p = \pi(x),$$

thus establishing that (v_i) allow to represent the preferences as claimed. Since all $(\tilde{v}_{\pi,i})$ must be positive affine transformations of $(v_{\pi,i})$, which are continuous, both $(\tilde{v}_{\pi,i})$ and (v_i) also are families of continuous functions.

Step 3 Normalize (v_i, r_i) in relation to π^0.
Fix the scaling-invariant context π^0, which exists by Axiom 4, we know from Step 2

$$u_{\pi^0}(x) = \sum_{i \leq n} v_i\big(\pi_i^0(x) - r_i(\pi^0)\big), \tag{7}$$

which in turn implies that we can translate (v_i) and (r_i) such that

$$u_{\pi^0}(x) = \sum_{i \leq n} v_i\big(\pi_i^0(x)\big), \tag{8}$$

i.e. such that $r_i(\pi^0) = 0$ for all $i \leq n$. Again, by $r(\pi) + c = r(\pi + c)$ for all $c \in \mathbb{R}^n$, this implies $r(\pi) = c$ if $\pi = \pi^0 + c$, for all $\pi \in \Pi$. Note that given this translation, we can analyze narrow bracketing and broad bracketing in a uniform manner when focusing on π^0 (i.e. we do not have to include r_i as $r_i(\pi^0) = 0$).

Step 4 Using scaling invariance to fix the functional form.
By Axiom 4, preferences in context π^0 are scaling invariant. That is, for all $\lambda > 0$, define $u_{\lambda\pi^0} : X \to \mathbb{R}$ such that

$$u_{\lambda\pi^0}(x) = \sum_{i \leq n} v_i\big(\lambda \pi_i^0(x)\big), \tag{9}$$

for all λ, x, and we obtain

$$u_{\lambda\pi^0}(x) \geq u_{\lambda\pi^0}(y) \Leftrightarrow \qquad u_{\pi^0}(x) \geq u_{\pi^0}(y) \qquad \Leftrightarrow \qquad \pi^0(x) \succsim_{\pi^0} \pi^0(y). \tag{10}$$

By aforementioned Theorem III.4.1 of Wakker (1989) this implies that $u_{\lambda\pi^0}$ is a positive affine transformation of u_{π^0}, i.e. there exist $a : \mathbb{R}_+ \to \mathbb{R}$ and $b : \mathbb{R}_+ \to \mathbb{R}^n$ such that

$$v_i\big(\lambda \pi_i^0(x)\big) = v_i\big(\pi_i^0(x)\big) \cdot a(\lambda) + b_i(\lambda) \tag{11}$$

for all $i \in N$, $x \in X$, $\lambda > 0$. Now, define $X_i^+ = \{x \in X \,|\, \pi_i^0(x) > 0\}$ as well as $\tilde{\lambda} = \log \lambda$, $\tilde{v}_i : \mathbb{R} \to \mathbb{R}$ such that $\tilde{v}_i(\log p) = v_i(p)$ for all $p > 0$, and $\tilde{\pi}_i^0(x) = \log \pi_i^0(x)$ for all $x \in X_i^+$, which yields

$$\tilde{v}_i\big(\tilde{\lambda} + \tilde{\pi}_i^0(x)\big) = \tilde{v}_i\big(\tilde{\pi}_i^0(x)\big) \cdot a(\tilde{\lambda}) + b_i(\tilde{\lambda}). \tag{12}$$

By continuity of v_i we obtain continuity of $\tilde{v}_i$, and since the payoff image $\pi^0[X]$ is a cone in $\mathbb{R}^n$ with all dimensions being essential, it has positive volume in $\mathbb{R}^n$, i.e. $\pi_i^0[X]$ is an interval of positive length for all dimensions i. Hence, Theorem 1 and Corollary 1 of Aczél (1966, p. 150) imply that all solutions of this (Pexider) functional equation satisfy either

$$\tilde{v}_i\big(\tilde{\pi}_i^0(x)\big) = \alpha \cdot \tilde{\pi}_i^0(x) + \gamma \qquad \text{or} \qquad \tilde{v}_i\big(\tilde{\pi}_i^0(x)\big) = \alpha \cdot e^{\beta \tilde{\pi}_i^0(x)} + \gamma$$

with $\alpha \neq 0$ and β, γ being arbitrary constants, and inverting the variable substitutions,

$$v_i(\pi_i^0(x)) = \alpha \cdot \log \pi_i^0(x) + \gamma \qquad \text{or} \qquad v_i(\pi_i^0(x)) = \alpha \cdot \big(\pi_i^0(x)\big)^{\beta} + \gamma.$$

To distinguish the constants from constants in other dimensions, we rewrite

$$v_i(\pi_i^0(x)) = \alpha_i^+ + \beta_i^+ \cdot \log \pi_i^0(x) \qquad \text{or} \qquad v_i(\pi_i^0(x)) = \alpha_i^+ \cdot \big(\pi_i^0(x)\big)^{\beta_i^+} + \gamma_i^+$$

for all $x \in X_i^+$. Next, define $X_i^- = \{x \in X \,|\, \pi_i^0(x) < 0\}$, and apply the same line of arguments to $-\pi_i^0(x)$ for all $x \in X_i^-$, which yields

$$v_i(\pi_i^0(x)) = \alpha_i^- + \beta_i^- \cdot \log\big(-\pi_i^0(x)\big) \qquad \text{or} \qquad v_i(\pi_i^0(x)) = -\alpha_i^- \cdot \big(-\pi_i^0(x)\big)^{\beta_i^-} + \gamma_i^-$$

for all $x \in X_i^-$, again with $\alpha_i^- \neq 0$ and β_i^-, γ_i^- being arbitrary constants.

Step 5 Using continuity and Eq. (11) to normalize the parameters.
In the following, we refer to the two possible forms of the value function v_i as power form and logarithmic form (in the obvious manner). By continuity, the logarithmic form is feasible only if $\pi_i(x) > 0$ for all $x \in X$, implying that the second branch is never taken. Hence, for all $i \leq n$ and all $x \in X$,

$$v_i(\pi_i^0(x)) = \alpha_i^+ + \beta_i^+ \cdot \log\big(\pi_i^0(x)\big),$$

and we can set $\alpha_i^+ = 0$ for all i by applying a positive affine transformation (recalling that the value functions are unique up to positive affine transformation). This establishes the claim for the logarithmic form in context π^0, noting that α_i^+ and β_i^+ are switched (for the logarithmic form) in the formulation of the proposition for notational convenience.

Regarding the power form of the value function, rescaling payoffs we obtain

$$\begin{aligned}
&\forall x \in X_i^+ : \; v_i(\lambda \pi_i^0(x)) = \alpha_i^+ \cdot \big(\lambda \pi_i^0(x)\big)^{\beta_i^+} + \gamma_i^+ = \alpha_i^+ \cdot \big(\pi_i^0(x)\big)^{\beta_i^+} \cdot \lambda^{\beta_i^+} + \gamma_i^+ \\
&\forall x \in X_i^- : \; v_i(\lambda \pi_i^0(x)) = -\alpha_i^- \cdot \big(-\lambda \pi_i^0(x)\big)^{\beta_i^-} + \gamma_i^- = -\alpha_i^- \cdot \big(-\pi_i^0(x)\big)^{\beta_i^-} \cdot \lambda^{\beta_i^-} + \gamma_i^-,
\end{aligned}$$

which is compatible with Eq. (11) only if $\beta_i^+ = \beta_i^- = \beta$ and $\gamma_i^+ = \gamma_i^- = \gamma_i$ for all i. Given the latter, we can again set $\gamma_i^+ = \gamma_i^- = 0$ by a positive affine transformation. As a result, the claim for both the logarithmic form and the power form is established for context π^0.

Step 6 Extension to contexts $\pi \neq \pi^0$.
Narrow bracketing: Fix any $\pi \in \Pi$, and let $c \in \mathbb{R}^n$ such that $\pi = \pi^0 + c$. By Step 2, we know that

$$u_\pi(x) = \sum_{i \leq n} v_i\big(\pi_i(x) - r_i(\pi)\big)$$

represents $\succsim_\pi$ with v_i as characterized in the previous step and r_i as characterized in Steps 2 and 3. Since the representation is unique up to positive affine transformation, we can add arbitrary constants, and the claim is established for any context $\pi \in \Pi$.

Broad bracketing: By Step 2, the utility representation characterized in Step 5 applies uniformly to all contexts. □

A.2 Proofs of Propositions 2 and 3

A.2.1 Optimal choice of a regular dictator Δ in a given game Γ with $P_1 = [0,B]$

Note that since for this part of the proof the game Γ is kept fixed, we drop the game index on the utility function and write r_i instead of $r_i(\Gamma)$ for the reference points. Then dictator Δ's utility function in game Γ is given by

$$u(p_1) = \frac{1}{\beta} \times \begin{Bmatrix} (p_1 - r_1)^\beta & \text{if } p_1 \geq r_1 \\ -\delta(r_1 - p_1)^\beta & \text{if } p_1 < r_1 \end{Bmatrix} + \frac{\alpha}{\beta} \times \begin{Bmatrix} (p_2(p_1) - r_2)^\beta & \text{if } p_2(p_1) \geq r_2 \\ -\delta(r_2 - p_2(p_1))^\beta & \text{if } p_2(p_1) < r_2 \end{Bmatrix}$$

where $p_2(p_1) = t(B - p_1)$.

Step 1 Dictator Δ never chooses p_1 such that $p_1 < r_1$ and $p_2(p_1) < r_2$.

By satisfiability of reference points and $P_1 = [0,B]$ dictator Δ can always choose $p_1 \in P_1$ such that $p_1 \geq r_1$ and $p_2(p_1) \geq r_2$. This yields utility $u(p_1) = (p_1 - r_1)^\beta/\beta + \alpha(t(B-p_1) - r_2)^\beta/\beta \geq 0$ where the inequality follows by weak efficiency concerns ($0 < \beta < 1$). Choosing $p_1' \in P_1$ such that $p_1' < r_1$ and $p_2(p_1') < r_2$ instead yields utility $u(p_1') = -\delta(r_1 - p_1)^\beta/\beta - \alpha\delta(r_2 - t(B-p_1))^\beta/\beta < 0$.

Thus, we can restrict attention to the regions where at most one of the two players is in the loss-domain, i.e. does not reach her reference point. In the following we will first determine the local optima for dictator Δ in each of the three remaining regions. Then we can determine the global optimum by comparing utilities of the local optima.

Step 2 Local optimum in region 1: $p_1 \in [r_1, B - \frac{1}{t} r_2]$ ($\Leftrightarrow p_1 \geq r_1$ and $p_2(p_1) \geq r_2$)

The utility function that applies is

$$u^{(1)}(p_1) = (p_1 - r_1)^\beta + \alpha \cdot (t(B - p_1) - r_2)^\beta$$

Differentiating $u^{(1)}$ with respect to p_1 we get

$$\frac{du^{(1)}}{dp_1} = \beta (p_1 - r_1)^{\beta-1} - \alpha\beta t \,(t\,(B - p_1) - r_2)^{\beta-1}$$

which yields the first order condition

$$(p_1 - r_1)^{1-\beta} (t\,(B-p_1) - r_2)^{\beta-1} = \frac{1}{\alpha t} \qquad \Leftrightarrow \qquad \frac{t\,(B-p_1) - r_2}{p_1 - r_1} = (\alpha t)^{\frac{1}{1-\beta}}.$$

and the solution

$$p_1^+(\Gamma) = \frac{B + c_\alpha r_1 - r_2/t}{c_\alpha + 1} \quad \text{and} \quad p_2^+(\Gamma) = \frac{tc_\alpha\,(B - r_1) + r_2}{c_\alpha + 1}$$

using $c_\alpha := (\alpha t^\beta)^{\frac{1}{1-\beta}}$. Note that for $p_1 = B - \frac{1}{t} r_2$ and $p_1 = r_1$ the above first order condition is not defined because the utility function exhibits kinks at these points. We have $p_1^+(\Gamma) = B - \frac{1}{t} r_2 = r_1$

iff satisfiability is binding, i.e. $B - r_1 - \frac{1}{t}r_2 = 0$. By satisfiability we have $p_1^+(\Gamma) \in [r_1, B - \frac{1}{t}r_2]$ for all regular dictators Δ. Furthermore, the second order condition for $p_1^+(\Gamma)$ to be a maximum reduces to

$$t^{2-\beta}c_\alpha(1+c_\alpha)^{1-\beta}(t(B-r_1)-r_2)^\beta > 0,$$

which is fulfilled for $p_1^+(\Gamma)$ by satisfiability, weak efficiency concerns ($0 < \beta < 1$), and $\alpha, t > 0$. Overall, we thus have for the local optimum in region 1

$$p_1^{(*)} = p_1^+(\Gamma).$$

Step 3 Local optimum in region 2: $p_1 \in (B - \frac{1}{t}r_2, B]$ ($\Leftrightarrow p_1 \geq r_1$ and $p_2 < r_2$)

The utility function that applies is

$$u^{(2)}(p_1) = (p_1 - r_1)^\beta - \delta\alpha \cdot (r_2 - t(B - p_1))^\beta$$

Differentiating $u^{(2)}$ with respect to p_1 we obtain

$$\frac{du^{(2)}}{dp_1} = \beta(p_1 - r_1)^{\beta-1} - \delta\alpha\beta t\,(r_2 - t\,(B - p_1))^{\beta-1}$$

which yields the first order condition

$$(p_1 - r_1)^{1-\beta}(r_2 - t(B - p_1))^{\beta-1} = \frac{1}{\delta\alpha t} \qquad \Leftrightarrow \qquad \frac{r_2 - t(B-p_1)}{p_1 - r_1} = (\delta\alpha t)^{\frac{1}{1-\beta}}$$

and the solution

$$p_1^{(2)}(\Gamma) = \frac{B - \delta^{\frac{1}{1-\beta}}c_\alpha r_1 - r_2/t}{1 - \delta^{\frac{1}{1-\beta}}c_\alpha} \quad \text{and} \quad p_2^{(2)}(\Gamma) = \frac{t\delta^{\frac{1}{1-\beta}}c_\alpha(r_1 - B) + r_2}{1 - \delta^{\frac{1}{1-\beta}}c_\alpha}.$$

By satisfiability we have $p_1^{(2)} \in (B - \frac{1}{t}r_2, B]$ iff $\delta^{\frac{1}{1-\beta}}c_\alpha \leq \frac{r_2}{t(B-r_1)} \Leftrightarrow \delta \leq \frac{1}{\alpha t^\beta}\left(\frac{r_2}{t(B-r_1)}\right)^{1-\beta}$. Using $\delta^{\frac{1}{1-\beta}}c_\alpha < 1$, the second order condition for $p_1^{(2)}(\Gamma)$ to be a maximum reduces to

$$t^{2-\beta}\delta^{\frac{1}{1-\beta}}c_\alpha(1 - \delta^{\frac{1}{1-\beta}}c_\alpha)^{1-\beta}\,(t(B-r_1)-r_2)^\beta < 0.$$

Thus, the second order condition does not hold for any $p_1^{(2)}(\Gamma) \in (B - \frac{1}{t}r_2, B]$ by satisfiability and weak efficiency concerns ($0 < \beta < 1$). It follows that the local optimum is either $p_1 = B - \frac{1}{t}r_2$ or $p_1 = B$ depending on whether $u^{(2)}(B - \frac{1}{t}r_2) \geq u^{(2)}(B)$, a condition which reduces to

$$\delta \geq c_\alpha^{\beta-1}\left(\left(\frac{t(B-r_1)}{r_2}\right)^\beta - \left(\frac{t(B-r_1)}{r_2} - 1\right)^\beta\right).$$

Overall, we thus have for the local optimum in region 2

$$p_1^{(*)} = \begin{cases} B - \frac{1}{t}r_2 & \text{if } \delta \geq c_\alpha^{\beta-1}\left(\left(\frac{t(B-r_1)}{r_2}\right)^\beta - \left(\frac{t(B-r_1)}{r_2} - 1\right)^\beta\right) \\ B & \text{else.} \end{cases}$$

Step 4 Local optimum in region 3: $p_1 \in [0, r_1)$ ($\Leftrightarrow p_1 < r_1$ and $p_2(p_1) \geq r_2$)

The utility function that applies is

$$u^{(3)}(p_1) = -\delta \cdot (r_1 - p_1)^\beta + \alpha \cdot (t(B - p_1) - r_2)^\beta$$

Differentiating $u^{(3)}$ with respect to p_1 we obtain

$$\frac{du^{(3)}}{dp_1} = \delta\beta(r_1 - p_1)^{\beta-1} - \alpha\beta t(t(B - p_1) - r_2)^{\beta-1}$$

which yields the first order condition

$$(r_1 - p_1)^{1-\beta}(t(B - p_1) - r_2)^{\beta-1} = \frac{\delta}{\alpha t} \qquad \Leftrightarrow \qquad \frac{t(B - p_1) - r_2}{r_1 - p_1} = \left(\frac{\alpha t}{\delta}\right)^{\frac{1}{1-\beta}}$$

and the solution

$$p_1^{(3)}(\Gamma) = \frac{B - \delta^{1-\beta}c_\alpha r_1 - r_2/t}{1 - \delta^{1-\beta}c_\alpha} \quad \text{and} \quad p_2^{(3)}(\Gamma) = \frac{t\delta^{1-\beta}c_\alpha(r_1 - B) + r_2}{1 - \delta^{1-\beta}c_\alpha}$$

By satisfiability we have $p_1^{(3)} \in [0, r_1)$ iff $\delta^{1-\beta}c_\alpha \geq \frac{tB-r_2}{tr_1} \Leftrightarrow \delta \leq \alpha t^\beta \left(\frac{tr_1}{tB-r_2}\right)^{1-\beta}$. The second order condition for $p_1^{(3)}(\Gamma)$ to be a maximum reduces to

$$\frac{1}{\delta^{1-\beta}c_\alpha}\left(\frac{r_1 - B + r_2/t}{1 - \delta^{1-\beta}c_\alpha}\right)^{\beta-2} > \left(\frac{r_1 - B + r_2/t}{1 - \delta^{1-\beta}c_\alpha}\right)^{\beta-2},$$

which by satisfiability does not hold for any $p_1^{(3)}(\Gamma) \in [0, r_1)$. It follows that the optimum is either $p_1 = 0$ or $p_1 = r_1$ depending on whether $u^{(3)}(0) \geq u^{(3)}(r_1)$, a condition which reduces to

$$\delta \leq c_\alpha^{1-\beta}\left(\left(\frac{tB - r_2}{tr_1}\right)^\beta - \left(\frac{tB - r_2}{tr_1} - 1\right)^\beta\right).$$

Overall, we thus have for the local optimum in region 3

$$p_1^{(*)} = \begin{cases} 0 & \text{if } \delta \leq c_\alpha^{1-\beta}\left(\left(\frac{tB-r_2}{tr_1}\right)^\beta - \left(\frac{tB-r_2}{tr_1} - 1\right)^\beta\right) \\ r_1 & \text{else.} \end{cases}$$

Step 5 Reducing the set of candidate solutions for the global optimum

Using weak efficiency concerns ($0 < \beta < 1$) and $\alpha, t \geq 0$ we have $u(p_1^+(\Gamma)) \geq u(B - \frac{1}{t}r_2)$ and $u(p_1^+(\Gamma)) \geq u(r_1)$ for all regular dictators Δ, a result which obtains by simple rearrangement of the two inequalities. Thus, the remaining candidate solutions for the overall utility maximizer are $p_1 = p_1^+(\Gamma)$, $p_1 = B$, and $p_1 = 0$.

Furthermore, we have $u(p_1^+(\Gamma)) \geq u(0)$ iff

$$\delta \geq c_\alpha^{1-\beta}\left(\frac{tB - r_2}{tr_1}\right)^\beta - (c_\alpha + 1)^{1-\beta}\left(\frac{tB - r_2}{tr_1} - 1\right)^\beta. \tag{13}$$

From weak efficiency concerns ($0 < \beta < 1$) we can conclude that

$$c_\alpha^{1-\beta} < (c_\alpha + 1)^{1-\beta}.$$

Define $f(x) = x^\beta$, then weak efficiency concerns ($0 < \beta < 1$) imply that f is subadditive in the domain $\mathbb{R}^+$, i.e. $f(a) + f(b) \geq f(a + b) \forall a, b \geq 0$. Thus, using satisfiability and letting $a =$

$\frac{tB-r_2}{tr_1}-1$ and $b=1$, we have

$$f(a)+f(b)=\left(\frac{tB-r_2}{tr_1}-1\right)^{\beta}+1^{\beta}\geq\left(\frac{tB-r_2}{tr_1}\right)^{\beta}=f(a+b)$$

implying

$$\left(\frac{tB-r_2}{tr_1}\right)^{\beta}-\left(\frac{tB-r_2}{tr_1}-1\right)^{\beta}\leq 1.$$

Suppose $c_{\alpha}^{1-\beta}\leq 1$. In this case we can conclude that the lower bound for δ defined in (13) is lower or equal 1, which by weak loss aversion ($\delta\geq 1$) implies $u(p_1^{+}(\Gamma))\geq u(0)$. Note that $c_{\alpha}^{1-\beta}\leq 1$ by weak altruism ($0\leq\alpha\leq 1$) always holds under no efficiency gains from giving ($t\leq 1$) such that in this case the candidate solutions for the overall utility maximizer reduce further to $p_1=p_1^{+}(\Gamma)$ and $p_1=B$.

Finally, we have $u(p_1^{+}(\Gamma))\geq u(B)$ iff

$$\delta\geq c_{\alpha}^{\beta-1}\left(\left(\frac{t(B-r_1)}{r_2}\right)^{\beta}-(c_{\alpha}+1)^{1-\beta}\left(\frac{t(B-r_1)}{r_2}-1\right)^{\beta}\right). \tag{14}$$

Suppose $c_{\alpha}^{1-\beta}>1$. In this case by a similar argument as above we can conclude that the lower bound for δ defined in (14) is lower or equal 1, which by weak loss aversion ($\delta\geq 1$) implies $u(p_1^{+}(\Gamma))\geq u(B)$. We can therefore conclude that under efficiency gains from giving ($t>1$) the candidate solutions for the overall utility maximizer reduce to $p_1=p_1^{+}(\Gamma)$ and $p_1=B$ in case $c_{\alpha}^{1-\beta}\leq 1$ while they reduce to $p_1=p_1^{+}(\Gamma)$ and $p_1=0$ in case $c_{\alpha}^{1-\beta}>1$.

Step 6 Global optimum

For the global optimum we have to distinguish the following two cases:

- Case 1: $c_{\alpha}^{1-\beta}\leq 1$

  $$p_1^{*}=\begin{cases}p_1^{+}(\Gamma) & \text{if } \delta\geq\delta^{+}(\Gamma)\\ B & \text{else.}\end{cases}$$

  with

  $$\delta^{+}(\Gamma):=c_{\alpha}^{\beta-1}\left(\left(\frac{t(B-r_1)}{r_2}\right)^{\beta}-(c_{\alpha}+1)^{1-\beta}\left(\frac{t(B-r_1)}{r_2}-1\right)^{\beta}\right)$$

- Case 2: $c_{\alpha}^{1-\beta}>1$

  $$p_1^{*}=\begin{cases}p_1^{+}(\Gamma) & \text{if } \delta\geq\delta^{-}(\Gamma)\\ 0 & \text{else.}\end{cases}$$

  with

  $$\delta^{-}(\Gamma):=c_{\alpha}^{1-\beta}\left(\frac{tB-r_2}{tr_1}\right)^{\beta}-(c_{\alpha}+1)^{1-\beta}\left(\frac{tB-r_2}{tr_1}-1\right)^{\beta}$$

Note that under no efficiency gains from giving ($t\leq 1$) only case 1 applies.

A.2.2 Establishing the comparative statics

Step 1 Non-convexity In any game Γ with $P_1 = [0, B]$ there are dictators with non-convex preferences.

Fix a game Γ with $P_1 = [0, B]$. Consider a dictator Δ with $\delta \leq \bar{\delta}(\Gamma)$ where

$$\bar{\delta}(\Gamma) := c_\alpha^{\beta-1} \left(\frac{r_2(\Gamma)}{t(B - r_1(\Gamma))} \right)^{1-\beta} .$$

We have shown in step 3 of A.2.1 that the utility function of this dictator attains a minimum at $p_1 = p_1^{(2)}(\Gamma) \in [B - r_2/t, B]$ and has no other local extrema in that region. Furthermore, we have shown in step 2 of A.2.1 that her utility function attains a maximum at $p_1 = p_1^+(\Gamma) \in [r_1, B - r_2/t]$ and has no other local extrema in that region. Consider options a and b with $p_1^a = B$ and $p_1^b = p_1^+(\Gamma)$. Construct option c by choosing $\lambda \in [0,1]$ such that $p_1^c = \lambda p_1^a + (1-\lambda) p_1^b = p_1^{(2)}(\Gamma)$. Then, for dictator Δ in game Γ there exists an option d with $p_1^d \in (p_1^+(\Gamma), B)$ such that $u_\Gamma(p_1^a) \geq u_\Gamma(p_1^d)$ and $u_\Gamma(p_1^b) \geq u_\Gamma(p_1^d)$ but $u_\Gamma(p_1^c) < u_\Gamma(p_1^d)$. Since u_Γ represents dictator Δ's preferences in game Γ, this implies that her preferences are non-convex.

We still have to show that in any game Γ with $P_1 = [0, B]$ there exist regular dictators with $\delta \leq \bar{\delta}(\Gamma)$. For any transfer rate t specified by Γ we can find (α, β) satisfying $0 \leq \alpha \leq 1$ and $0 < \beta < 1$ such that $c_\alpha^{1-\beta} \leq 1 \Leftrightarrow c_\alpha^{\beta-1} \geq 1$. Given such (α, β), for any endowments (B_1, B_2) specified by Γ, we can find (w_1, w_2) in accordance with satisfiability resulting in reference points $r_1(\Gamma) = w_1 B_1 + w_2 B_2$ and $r_2(\Gamma) = t(w_1 B_2 + w_2 B_1)$ such that $r_2(\Gamma)/t(B - r_1(\Gamma))$ is close enough to 1 to make $\bar{\delta}(\Gamma) \geq 1$. Thus, given such $(\alpha, \beta, w_1, w_2)$, we can conclude that there exist δ satisfying weak loss aversion ($\delta \geq 1$) such that $\delta \leq \bar{\delta}(\Gamma)$.

Step 2 Taking options reduce giving both at the extensive and intensive margin Introducing a taking option turns some initial givers into takers and reduces average amounts given.

Consider two games $\Gamma = \langle B_1, B_2, P_1, t \rangle$ and $\Gamma' = \langle B_1, B_2, P_1', t \rangle$ with $B_2 > 0$ that are equivalent in every dimension except the choice set of the dictator. In Γ the choice set is restricted to $P_1 = [0, \max p_1]$ with $\max p_1 = B_1$ and in Γ' the choice set is extended to $p_1' = [0, \max p_1']$ with $B_1 < \max p_1' \leq B_1 + B_2$.

Moving from Γ to Γ' the only game parameter that changes is the maximum payoff for the dictator which rises from $\max p_1 = B_1$ to $\max p_1'$. As a result of this rise, the minimum payoff for the recipient adjusts accordingly, i.e. it falls from $\min p_2 = t(B_1 + B_2 - \max p_1) = tB_2$ to $\min p_2' = t(B_1 + B_2 - \max p_1')$. Therefore, the utility functions of a regular dictator Δ in Γ and Γ' differ in the players' reference points. We have

$$r_2(\Gamma) = t\,(B_2 + (1 - w_1 + w_2)B_1 - (1 - w_1)\max p_1) \quad \text{with} \quad \frac{dr_2}{d \max p_1} = -t(1 - w_1) \leq 0,$$

and

$$r_1(\Gamma) = (w_1 - w_2)B_1 + w_2 \max p_1 \quad \text{with} \quad \frac{dr_1}{d \max p_1} = w_2 \geq 0,$$

where the inequalities follow from satisfiability. Thus, we have $r_2(\Gamma) \geq r_2(\Gamma')$ and $r_1(\Gamma) \leq r_1(\Gamma')$. Plugging in our reference points we get for the interior solution in game Γ

$$p_1^+(\Gamma) = (w_1 - w_2)B_1 + \frac{1 - w_1 + c_\alpha w_2}{c_\alpha + 1} \max p_1$$

and the derivative with respect to the maximum payoff of the dictator is given by

$$\frac{dp_1^+}{d\max p_1} = \frac{1-w_1+c_\alpha w_2}{c_\alpha+1} \geq 0$$

where the inequality follows from $\alpha \geq 0$ and satisfiability. Thus, we have $p_1^+(\Gamma) \leq p_1^+(\Gamma')$. Note furthermore, that by satisfiability $\frac{dp_1^+}{d\max p_1} \leq 1$ implying that the interior solution is feasible for any regular dictator in Γ and Γ'.

In A.2.1. we specified the global optimum for games like Γ with $P_1 = [0,B]$. In games like Γ' where the choice set of the dictator is restricted to $P_1' = [0, \max p_1]$ with $\max p_1 < B$ the selfish corner solution $p_1 = B$ is not feasible. Thus, we have for $c_\alpha^{1-\beta} \leq 1$ (case 1):

$$p_1^* = \begin{cases} p_1^+(\Gamma) & \text{if } \delta \geq \hat{\delta}^+(\Gamma) \\ \max p_1 & \text{else.} \end{cases}$$

with

$$\hat{\delta}^+(\Gamma) := c_\alpha^{\beta-1}\left(\left(\frac{t(\max p_1 - r_1)}{r_2 - t(B-\max p_1)}\right)^\beta - (c_\alpha+1)^{1-\beta}\left(\frac{t(B-r_1)-r_2}{r_2-t(B-\max p_1)}\right)^\beta\right)$$

where the expression for $\hat{\delta}^+(\Gamma)$ follows from rearrangement of $u_\Gamma(p_1^+(\Gamma)) \geq u_\Gamma(\max p_1)$. Note that for $c_\alpha^{1-\beta} > 1$ (case 2) the specification of the global optimum is not affected by the restriction of the choice set because the altruistic corner solution $p_1 = 0$ is feasible in Γ'.

We consider this threshold $\hat{\delta}^+(\Gamma')$ such that in game Γ' among the regular dictators with $c_\alpha^{1-\beta} \leq 1$, those with $\delta < \hat{\delta}^+(\Gamma')$ choose the selfish corner solution $p_1 = \max p_1'$ while those with $\delta \geq \hat{\delta}^+(\Gamma')$ choose the interior solution $p_1 = p_1^+(\Gamma')$. We can rewrite it as

$$\hat{\delta}^+(\Gamma') := c_\alpha^{\beta-1}\left(\left(\frac{(1-w_2)\max p_1' - (w_1-w_2)B_1}{w_1\max p_1' - (w_1-w_2)B_1}\right)^\beta - (c_\alpha+1)^{1-\beta}\left(\frac{(1-w_2)\max p_1' - (w_1-w_2)B_1}{w_1\max p_1' - (w_1-w_2)B_1} - 1\right)^\beta\right).$$

Then the derivative with respect to $\max p_1'$ is given by

$$\frac{d\hat{\delta}^+}{d\max p_1'} = \frac{\beta(1-w_1-w_2)(w_1-w_2)B_1}{c_\alpha^{1-\beta}(w_1\max p_1' - (w_1-w_2)B_1)^2}\left((c_\alpha+1)^{1-\beta}\left(\frac{(1-w_2)\max p_1' - (w_1-w_2)B_1}{w_1\max p_1' - (w_1-w_2)B_1} - 1\right)^{\beta-1} - \left(\frac{(1-w_2)\max p_1' - (w_1-w_2)B_1}{w_1\max p_1' - (w_1-w_2)B_1}\right)^{\beta-1}\right).$$

From weak altruism, weak efficiency concerns, satisfiability, and $w_1 \geq w_2$ we can conclude that $\frac{d\hat{\delta}^+}{d\max p_1'} \geq 0$. Thus, we have $\hat{\delta}^+(\Gamma) \leq \hat{\delta}^+(\Gamma')$, implying that weakly more regular dictators with $c_\alpha^{1-\beta} \leq 1$ prefer the selfish corner solution to the interior solution in Γ' compared to Γ.

Now, consider the threshold $\delta^-(\Gamma)$ such that in game Γ among the regular dictators with $c_\alpha^{1-\beta} > 1$, those with $\delta < \delta^-(\Gamma)$ prefer the altruistic corner solution $p_1 = 0$ while those with $\delta \geq \delta^-(\Gamma)$ prefer the interior solution $p_1 = p_1^+(\Gamma)$. We can rewrite the threshold as

$$\delta^-(\Gamma) = c_\alpha^{1-\beta}\left(\frac{(1-w_1)\max p_1 + (w_1-w_2)B_1}{w_2\max p_1 + (w_1-w_2)B_1}\right)^\beta - (c_\alpha+1)^{1-\beta}\left(\frac{(1-w_1)\max p_1 + (w_1-w_2)B_1}{w_2\max p_1 + (w_1-w_2)B_1} - 1\right)^\beta.$$

Then the derivative with respect to $\max p_1$ is given by

$$\frac{d\delta^-}{d\max p_1} = \frac{\beta(1-w_1-w_2)(w_1-w_2)B_1}{(w_2\max p_1 + (w_1-w_2)B_1)^2}\left(c_\alpha^{1-\beta}\left(\frac{(1-w_1)\max p_1 + (w_1-w_2)B_1}{w_2\max p_1 + (w_1-w_2)B_1}\right)^{\beta-1} - (c_\alpha+1)^{1-\beta}\left(\frac{(1-w_1)\max p_1 + (w_1-w_2)B_1}{w_2\max p_1 + (w_1-w_2)B_1} - 1\right)^{\beta-1}\right).$$

From weak altruism, weak efficiency concerns, satisfiability, and $w_1 \geq w_2$ we can conclude that $\frac{d\delta^-}{d\max p_1} \leq 0$. Thus, we have $\delta^-(\Gamma) \geq \delta^-(\Gamma')$ implying that weakly less regular dictators with $c_\alpha^{1-\beta} > 1$ prefer the altruistic corner solution to the interior solution in Γ' compared to Γ.

Using these results together with our results from A.2.1 we can show that comparing the choice of any regular dictator Δ in Γ to her choice in Γ' one of the following cases applies:

(i) Her choice switches from $p_1 = p_1^+(\Gamma)$ to $p_1 = p_1^+(\Gamma')$ where $p_1^+(\Gamma) \leq p_1^+(\Gamma')$.

(ii) Her choice switches from $p_1 = p_1^+(\Gamma)$ to $p_1 = \max p'_1$ where $p_1^+(\Gamma) < \max p'_1$.

(iii) Her choice switches from $p_1 = 0$ to $p_1 = p_1^+(\Gamma')$ where $0 \leq p_1^+(\Gamma')$.

(iv) Her choice remains at $p_1 = 0$.

First, we restrict attention to regular dictators with $c_\alpha^{1-\beta} \leq 1$. Note that in game Γ by satisfiability $r_2(\Gamma) \leq B_2$ such that there is no feasible choice for the dictator in which the recipient's reference point is not fulfilled. Thus, in game Γ these dictators all choose the interior solution $p_1 = p_1^+(\Gamma)$. Now consider the same dictators in Γ' and split them into two groups according to their loss aversion parameters. The dictators with $\delta \geq \hat{\delta}^+(\Gamma')$ choose $p_1 = p_1^+(\Gamma')$ in Γ'. The dictators with $\delta < \hat{\delta}^+(\Gamma')$ choose $p_1 = \max p'_1$ in Γ'.

Now, restrict attention to regular dictators with $c_\alpha^{1-\beta} > 1$. We split these dictators into three groups according to their loss aversion parameters. Consider first the dictators with $\delta \geq \delta^-(\Gamma)$. These dictators choose $p_1 = p_1^+(\Gamma)$ in Γ. Since $\delta^-(\Gamma) \geq \delta^-(\Gamma')$ they choose $p_1 = p_1^+(\Gamma')$ in Γ'. Second, consider the dictators with $\delta \in [\delta^-(\Gamma'), \delta^-(\Gamma))$. These dictators choose $p_1 = 0$ in Γ and switch to $p_1 = p_1^+(\Gamma')$ in Γ'. Third, consider the dictators with $\delta < \delta^+(\Gamma')$. These dictators choose $p_1 = 0$ both in Γ and in Γ'.

We still have to show that for any Γ and Γ' there exist regular dictators who give in Γ and switch to taking in Γ'. We show that for any Γ and Γ' there exist regular dictators with $p_1^+(\Gamma) < B_1$ and $\delta < \hat{\delta}^+(\Gamma')$, i.e. regular dictators who give at the interior solution in Γ and to whom case (ii) applies. We have $p_1^+(\Gamma) < B_1$ iff $c_\alpha(1-w_1) + w_2 > 0$. Thus, we have $p_1^+(\Gamma) < B_1$ for all regular dictators with $0 < w_1 < 1$ or $w_1 = 1$ and $w_2 > 0$. Now, for any transfer rate t specified by Γ and Γ' we can find (α, β) satisfying $0 < \alpha \leq 1$ and $0 < \beta < 1$ such that $c_\alpha^{1-\beta} < 1 \Leftrightarrow c_\alpha^{\beta-1} > 1$. Given such (α, β), for any endowments and choice set (B_1, B_2, P'_1) specified by Γ' we have $((1-w_2)\max p'_1 - (w_1 - w_2)B_1)/(w_1 \max p'_1 - (w_1 - w_2)B_1) = 1$ for $w_1 = 1$ and $w_2 = 0$. Thus, by continuity of $\hat{\delta}^+$ we can always find $w_1 > 0$ and $w_2 \geq 0$ in accordance with satisfiability such that the expression is close enough to 1 to make $\hat{\delta}^+(\Gamma') > 1$ and given such $(\alpha, \beta, w_1, w_2)$, we can conclude that there exist δ satisfying weak loss aversion ($\delta \geq 1$) such that $\delta < \hat{\delta}^+(\Gamma')$.

Finally, we need to show that for any Γ and Γ' there exist regular dictators who give more in Γ than in Γ'. We show that for any Γ and Γ' there exist regular dictators with $p_1^+(\Gamma), p_1^+(\Gamma') < B_1$ and $\delta \geq \hat{\delta}^+(\Gamma')$. As above we have $p_1^+(\Gamma) < B_1$ for all regular dictators with $0 < w_1 < 1$ or $w_1 = 1$ and $w_2 > 0$. Furthermore, we have $\frac{dp_1^+}{d\max p_1} = 0$ for $w_1 = 1$ and $w_2 = 0$. Thus, by continuity of $p_1^+(\Gamma)$ for any transfer rate t specified by Γ and Γ' we can find $0 < w_1 \leq 1$ and $w_2 \geq 0$ in accordance with satisfiability such that $p_1^+(\Gamma) < p_1^+(\Gamma') < B_1$. Since there is no upper bound on the loss aversion parameter of regular dictators given such (w_1, w_2) there always exist regular dictators with $\delta \geq \hat{\delta}^+(\Gamma')$.

Step 3 Incomplete crowding out Reallocating initial endowment from dictator to recipient results (in expectation) in a payoff increase for the recipient.

Consider two games $\Gamma = \langle B_1, B_2, P_1, t\rangle$ and $\Gamma' = \langle B'_1, B'_2, P'_1, t\rangle$ without taking option, i.e. $P_1 = [0, B_1]$ and $P'_1 = [0, B'_1]$, where Γ' is generated from Γ by reallocating initial endowment from the dictator to the recipient, i.e. $B_1 + B_2 = B'_1 + B'_2 = \bar{B}$ and $B_1 < B'_1$. Thus, comparing such games we can write the recipient's endowment as a function of the dictator's endowment, i.e. $B_2(B_1) = \bar{B} - B_1$.

Moving from Γ to Γ' the game parameters that change are the player's endowments and the maximum payoff for the dictator. The dictator's endowment falls from B_1 to B'_1 while the recipient's endowment rises from $\bar{B} - B_1$ to $\bar{B} - B'_1$. Furthermore, the maximum payoff for the dictator falls from B_1 to B_1' such that the minimum payoff for the recipient rises from $\min p_2 = t(\bar{B} - B_1)$ to $\min p_2' = t(\bar{B} - B_1')$. Therefore, the utility functions of a regular dictator Δ in Γ and Γ' differ in the reference points of the dictator and the recipient. We have

$$r_1(\Gamma) = w_1 B_1 \quad \text{with} \quad \frac{dr_1}{dB_1} = w_1 \geq 0,$$

where the inequality follows from satisfiability, and

$$r_2(\Gamma) = t(\bar{B} - (1 - w_2)B_1) \quad \text{with} \quad \frac{dr_2}{dB_1} = -t(1 - w_2) \leq 0,$$

where the inequality follows from satisfiability and $t > 0$. Thus, we have $r_1(\Gamma) \geq r_1(\Gamma')$ and $r_2(\Gamma) \leq r_2(\Gamma')$. We can rewrite the interior solution as

$$p_1^+(\Gamma) = \frac{1 + c_\alpha w_1 - w_2}{c_\alpha + 1} B_1.$$

Taking the derivative with respect to the dictator's initial endowment we get

$$\frac{dp_1^+}{dB_1} = \frac{1 + c_\alpha w_1 - w_2}{c_\alpha + 1} \geq 0$$

where the inequality follows from imperfect altruism, weak efficiency concerns, and satisfiability. Thus, we have $p_1^+(\Gamma) \geq p_1^+(\Gamma')$.

Consider now the threshold for $\delta^-(\Gamma)$ such that in game Γ among the regular dictators with $c_\alpha^{1-\beta} > 1$, those with $\delta < \delta^-(\Gamma)$ choose the altruistic corner solution $p_1 = 0$ while those with $\delta \geq \delta^-(\Gamma)$ choose the interior solution $p_1 = p_1^+(\Gamma)$. We can rewrite the threshold as

$$\delta^-(\Gamma) = c_\alpha^{1-\beta} \left(\frac{1 - w_2}{w_1} \right)^\beta - (c_\alpha + 1)^{1-\beta} \left(\frac{1 - w_2}{w_1} - 1 \right)^\beta$$

and since the threshold is independent of B_1 we get $\frac{d\delta^-}{dB_1} = 0$. Thus, we have $\delta^-(\Gamma) = \delta^-(\Gamma') =: \delta^-$.

Using these results together with our results from A.2.1 we can show that comparing the choice of any regular dictator Δ in Γ to her choice in Γ' one of the following cases applies:

(i) Her choice switches from $p_1 = p_1^+(\Gamma)$ to $p_1 = p_1^+(\Gamma')$ where $p_1^+(\Gamma) \geq p_1^+(\Gamma')$.

(ii) Her choice remains at $p_1 = 0$.

Consider first only regular dictators with $c_\alpha^{1-\beta} \leq 1$. Since in neither Γ nor Γ' there is a feasible choice such that the reference point of the recipient is not fulfilled, these dictators all choose the respective interior solution in Γ and Γ'.

Now, consider regular dictators with $c_\alpha^{1-\beta} > 1$. We split these dictators into two groups according to their loss aversion parameters. Consider first the dictators with $\delta \geq \delta^-$. These dictators choose $p_1 = p_1^+(\Gamma)$ in Γ and $p_1 = p_1^+(\Gamma')$ in Γ'. Second, consider the dictators with $\delta < \delta^-$. These dictators choose $p_1 = 0$ both in Γ and Γ'.

Finally, we show that for any Γ and Γ' there exist regular dictators to whom case (i) applies in a strict sense, i.e. regular dictators whose choice in Γ' compared to Γ strictly increases the payoff of the recipient. For any transfer rate t specified by Γ and Γ' we can find $\alpha > 0$ and β satisfying weak altruism and weak efficiency concerns such that $c_\alpha^{1-\beta} \leq 1$, i.e. for any transfer rate t we can find regular dictators to whom case (i) applies. Furthermore, given such (α, β) we can always find (w_1, w_2) in accordance with satisfiability such that $dp_1^+/dB_1 > 0$.

Step 4 Efficiency concerns The recipient's payoff is weakly increasing in the transfer rate.

Consider two games $\Gamma = \langle B_1, B_2, P_1, t\rangle$ and $\Gamma' = \langle B_1, B_2, P_1, t'\rangle$ with $t < t'$, $P_1 = [0, \max p_1]$, and $B_1 \leq \max p_1 \leq B_1 + B_2$ which are equivalent in every dimension except the transfer rate.

The utility functions of a regular dictator Δ in Γ and Γ' differ only in the reference points of the recipient. His endowment is multiplied with t' instead of t and his minimal payoff increases from $\min p_2 = t(B - \max p_1)$ to $\min p_2' = t'(B - \max p_1)$. We have

$$r_2(\Gamma) = t(B_2 + (1 - w_1 + w_2)B_1 - (1 - w_1)\max p_1)$$

with

$$\frac{dr_2}{dt} = B_2 + (1 - w_1 + w_2)B_1 - (1 - w_1)\max p_1 \geq 0$$

where the inequality follows by satisfiability and $\max p_1 \leq B_1 + B_2$. Thus, we have $r_2(\Gamma) \leq r_2(\Gamma')$. We can rewrite the interior solution as

$$p_1^+(\Gamma) = \frac{t((1 - w_1)\max p_1 + (w_1 - w_2)B_1) + (\alpha t)^{\frac{1}{1-\beta}} r_1(\Gamma)}{(\alpha t)^{\frac{1}{1-\beta}} + t}.$$

Taking the derivative with respect to the transfer rate we get

$$\frac{dp_1^+}{dt} = \frac{tc_\alpha\beta}{1-\beta}(r_1(\Gamma) - (1 - w_1)\max p_1 - (w_1 - w_2)B_1) = \frac{tc_\alpha\beta}{1-\beta}(w_1 + w_2 - 1)\max p_1 \leq 0$$

where the inequality follows from weak altruism, weak efficiency concerns, and satisfiability. Thus, we have $p_1^+(\Gamma) \geq p_1^+(\Gamma')$.

Consider now the threshold $\hat{\delta}^+(\Gamma)$ such that in a game Γ with $\max p_1 > B_1$ among the regular dictators with $c_\alpha^{1-\beta} \leq 1$, those with $\delta < \hat{\delta}^+(\Gamma)$ choose the selfish corner solution $p_1 = \max p_1$ while those with $\delta \geq \hat{\delta}^+(\Gamma)$ choose the interior solution $p_1 = p_1^+(\Gamma)$. We can rewrite this threshold as

$$\hat{\delta}^+(\Gamma) = \frac{1}{\alpha t^\beta}\left(\left(\frac{\max p_1 - r_1}{w_1 \max p_1 - (w_1 - w_2)B_1}\right)^\beta - \left(\left(\alpha t^\beta\right)^{\frac{1}{1-\beta}} + 1\right)^{1-\beta}\left(\frac{\max p_1 - r_1}{w_1 \max p_1 - (w_1 - w_2)B_1} - 1\right)^\beta\right).$$

Taking the derivative with respect to t we get

$$\frac{d\hat{\delta}^+}{dt} = \frac{\beta}{tc_\alpha^{1-\beta}}\left((c_\alpha + 1)^{-\beta}\left(\frac{\max p_1 - r_1}{w_1 \max p_1 - (w_1 - w_2)B_1} - 1\right)^\beta - \left(\frac{\max p_1 - r_1}{w_1 \max p_1 - (w_1 - w_2)B_1}\right)^\beta\right) \leq 0,$$

where the inequality follows from weak altruism, weak efficiency concerns, and satisfiability. Thus, we have $\hat{\delta}^+(\Gamma) \geq \hat{\delta}^+(\Gamma')$, implying that weakly more regular dictators with $c_\alpha^{1-\beta} \leq 1$ choose the selfish corner solution in Γ compared to Γ'.

Consider now the threshold $\delta^-(\Gamma)$ such that in game Γ among the regular dictators with $\alpha t^\beta > 1$, those with $\delta < \delta^-(\Gamma)$ choose the altruistic corner solution $p_1 = 0$ while those with $\delta \geq \delta^-(\Gamma)$ choose the interior solution $p_1 = p_1^+(\Gamma)$. We can rewrite this threshold as

$$\delta^-(\Gamma) = \alpha t^\beta \left(\frac{(1 - w_1)\max p_1 + (w_1 - w_2)B_1}{r_1}\right)^\beta - \left(\left(\alpha t^\beta\right)^{\frac{1}{1-\beta}} + 1\right)^{1-\beta}\left(\frac{(1 - w_1)\max p_1 + (w_1 - w_2)B_1}{r_1} - 1\right)^\beta.$$

Taking the derivative with respect to t we get

$$\frac{d\delta^-}{dt} = \frac{\beta}{t}\left(c_\alpha^{1-\beta}\left(\frac{B_1 - (1 - w_2)(\max p_1 - B_1)}{r_1}\right)^\beta - \frac{c_\alpha}{(c_\alpha + 1)^\beta}\left(\frac{B_1 - (1 - w_2)(\max p_1 - B_1)}{r_1} - 1\right)^\beta\right).$$

From weak altruism, weak efficiency concerns, and satisfiability we can conclude that $\frac{d\delta^-}{dt} \geq 0$. Thus, we have $\delta^-(\Gamma) \leq \delta^-(\Gamma')$, implying that weakly more regular dictators with $\alpha t^\beta > 1$ choose the altruistic corner solution in Γ' compared to Γ.

Step 5 Reluctant sharers When an outside option is introduced, some initial givers switch to that option while the behavior of dictators who sort into the game stays unaffected.

Consider two games $\Gamma = \langle B_1, B_2, P_1, t\rangle$ and $\Gamma' = \langle B_1, B_2, P_1', t\rangle$ with $B_1 > 0$, $B_2 = 0$, $P_1 = [0, B_1]$, and $P_1' = \{[0, B_1], \tilde{p}_1\}$ where $0.5B_1 < \tilde{p}_1 \leq B_1$, i.e. game Γ' is generated from game Γ by adding an outside option to the choice set of the dictator.

Since the two games differ only in the choice set of the dictator, which is equivalent in both games except for the extra outside option in game Γ', the utility functions of a regular dictator in Γ and Γ' are equivalent where the two choice sets overlap. Furthermore, since the dictator's information is not manipulated by the choice of the outside option, her reference point stays the same for the choice of the outside option. We have $r_1(\Gamma) = r_1(\Gamma') =: r_1$ with $r_1 = w_1 B_1$. However, since the outside option leaves the recipient completely uninformed about the choice of the dictator and the rules of the game, his reference point is zero for the outside option choice. We thus have for the reference point of the recipient $r_2(\Gamma) = r_2(\Gamma') =: r_2$ with

$$r_2 = \begin{cases} t w_2 B_1 & \text{if } p_1 \in [0, B_1] \\ 0 & \text{if } p_1 = \tilde{p}_1 \end{cases}$$

The utility of a regular dictator if she chooses the outside option is then given by

$$u(\tilde{p}_1) = \begin{cases} \frac{1}{\beta}(\tilde{p}_1 - w_1 B_1)^\beta & \text{if } \tilde{p}_1 \geq w_1 B_1 \\ -\frac{\delta}{\beta}(w_1 B_1 - \tilde{p}_1)^\beta & \text{if } \tilde{p}_1 < w_1 B_1. \end{cases}$$

Since as noted above the utility functions of a regular dictator in Γ and Γ' are equivalent for $p_1 \in [0, B_1]$ we have $p_1^+(\Gamma) = p_1^+(\Gamma') =: p_1^+$ with

$$p_1^+ = \frac{1 + c_\alpha w_1 - w_2}{c_\alpha + 1} B_1.$$

and $\delta^+(\Gamma) = \delta^+(\Gamma') =: \delta^+$ with

$$\delta^+ = c_\alpha^{\beta-1}\left(\left(\frac{(1-w_1)B_1}{w_2}\right)^\beta - (c_\alpha + 1)^{1-\beta}\left(\frac{(1-w_1)B_1}{w_2} - 1\right)^\beta\right).$$

Note first, that no regular dictator with $w_1 > \tilde{p}_1/B_1$ chooses the outside option. By satisfiability, such a dictator can always choose $p_1 \in [0, B_1]$ such that $p_1 \geq r_1$ and $p_2(p_1) \geq r_2$. This yields utility $u(p_1) = (p_1 - r_1)^\beta/\beta + \alpha(t(B_1 - p_1) - r_2)^\beta/\beta \geq 0$, where the inequality follows from weak efficiency concerns. Choosing $p_1' = \tilde{p}_1$ instead yields $u(\tilde{p}_1) = -\delta(w_1 B_1 - \tilde{p}_1)^\beta < 0$. In the following we restrict attention to dictators with $w_1 \leq \tilde{p}_1/B_1$. We have $u(p_1^+) < u(\tilde{p}_1)$ iff

$$\tilde{p}_1 > B_1\left((c_\alpha + 1)^{\frac{1-\beta}{\beta}}(1 - w_1 - w_2) + w_1\right) =: \tilde{p}_1^{min}$$

We show that for any Γ and Γ' there exist regular dictators with $\delta \geq \delta^+$ and $\tilde{p}_1^{min} < \tilde{p}_1$, i.e. regular dictators who choose the interior solution in Γ and the outside option in Γ'. For any transfer rate t specified by Γ and Γ' we can find (α, β) satisfying weak altruism and weak efficiency concerns such that $c_\alpha^{1-\beta} \leq 1$. Given such (α, β), for any dictator endowment B_1 specified by Γ and Γ' and any outside option payment $\tilde{p}_1$ specified by Γ' we have $\tilde{p}_1^{min} = 0.5B_1$ for $w_1 = w_2 = 0.5$. Thus, by continuity of $\tilde{p}_1^{min}$ we can for any Γ and Γ' find (w_1, w_2) in accordance with satisfiability such that $\tilde{p}_1^{min} < B_1$. Since there is no upper bound on the loss aversion parameter of regular dictators, given such (w_1, w_2) there always exist regular dictators with $\delta \geq \delta^+$.

Step 6 Social pressure givers Ceteris paribus, higher susceptibility to social pressure implies higher recipient payoffs at the interior solution but also a higher propensity to choose the outside option in a sorting game.

Higher susceptibility to social pressure corresponds to a higher weight on the opponent's endowment in the reference points, i.e. a higher w_2. We have

$$\frac{\partial p_1^+}{\partial w_2} = -\frac{t}{c_\alpha + t} B_1 < 0 \quad \text{and} \quad \frac{\partial \tilde{p}_1^{min}}{\partial w_2} = -(c_\alpha + 1)^{\frac{1-\beta}{\beta}} B_1 \leq 0$$

where the inequalities follow from weak altruism and weak efficiency concerns.

□

B Details of the econometric specification

Technical details We estimate all parameters by maximum likelihood, and each case, the likelihood is maximized by a combination of two algorithms: first, using the robust (gradient-free) NEWUOA algorithm (Powell, 2006; Auger et al., 2009), secondly a Newton-Raphson method to ensure convergence. In addition, we cross-test globality of the maxima using a large number of informed starting values. These starting values are derived from estimates for related models on the same data set or from the same model on other data sets. Since we estimated the same model on many different data sets and related models on the same data sets, we were able to generate many informed starting vectors helpful in examining globality of maxima via cross-testing. As is well-known from numerical non-linear maximization (see e.g. McCullough and Vinod, 2003), generating informed starting values is necessary to ensure global optimality, and it proved extremely helpful also in our case. We stopped cross-testing and generating new starting values once the estimates had converged across all optimization problems simultaneously, based on which we conclude that we approximated the global maxima.

We evaluate significance of differences between models using the Schennach-Wilhelm likelihood ratio test (Schennach and Wilhelm, 2016). This test is robust to both misspecification and arbitrary nesting of models, which is required to allow for the possibility that all models are misspecified and to acknowledge that the nesting structure at least out-of-sample is not necessarily well-defined. In addition, the Schennach-Wilhelm test allows us cluster at the subject level and to thus account for the panel character of the data. We indicate significance of differences between models distinguishing the conventional level of 0.05 and the higher level of 0.01, which roughly implements the Bonferroni correction given four types of dictator game experiments we examine.

As many other experiments involving choice of numbers, responses in dictator games exhibit pronounced round-number patterns. We control for those using the focal choice adjusted logit model, exactly as derived and applied in Breitmoser (2017). The basic idea is that the roundedness of the number to be entered (to choose a given option) determine its “relative focality”, which is captured by a focality index $\phi : X \to \mathbb{R}$. The idea that focality is a choice-relevant attribute of options next to utility follows from Gul and Pesendorfer (2001), and given standard axioms including positivity, independence of irrelevant alternatives and narrow bracketing, this implies a generalized logit model of the form

$$\Pr(x) = \frac{\exp\{\lambda u(x) + \kappa \phi(x)\}}{\sum_{x'} \exp\{\lambda u(x') + \kappa \phi(x')\}}. \tag{15}$$

This approach effectively captures round-number effects in stochastic choice, and in turn, simply ignoring the round-number effects as pronounced as in Dictator games was shown to yield substantially biased results in Breitmoser (2017).[19] To avoid spending any degree of freedom here,

[19] For example, in the experiment of Korenok et al. (2014), subjects mostly picked multiples of five, typically from option sets ranging from 0 to 20. The most pronounced interior mass points are at choosing payoffs of 10 for both, dictator and recipient. Estimating the reference points of subjects in this experiments without controlling for round number effects yields estimates of reference point 10 each, and in this case, the reference point simply helps to capture the round-number effect. Controlling for the round-number effects, the overall model fit improves drastically and less round-number inspired reference points (deviating from 10 each) are estimated.

we use the same focality index as Breitmoser (2017)[20] and set κ equal to 0.8. Robustness checks on both choices are reported in Appendix C.

Capturing heterogeneity One of the more robust finding in behavioral economics is that subjects differ: They have heterogeneous preferences and differing precision in maximizing their preferences, and in addition, we suspect, they also have idiosyncratic reference points. Across subjects, these behavioral primitives are likely correlated. For example, a negative exponent β in the CES utility function implies a flat utility function, and thus to maintain "average precision" in maximizing utility a larger logit-parameter λ is required. Hence, β and λ generally are negatively correlated. For a related observation in the context of risk aversion, see for example Wilcox (2008). The correlation structure itself is unknown, however, and in addition, functional form assumptions about the marginal distributions of parameters seem to be equally difficult to make in the present context. We have only little knowledge about the distribution of individual preferences in generalized dictator games, except that the altruism weight α is likely truncated at say $(-0.5, 0.5)$, and that the exponent β does not seem to comply with a simple continuous distribution (for example, Andreoni and Miller, 2002, estimate that some subjects have linear preferences with β close to 1, some have Cobb-Douglas with $\beta \approx 0$, and others are Leontief with $\beta \to -\infty$).

While somewhat adequate approximations exist for each of these issues, we chose to tackle heterogeneity in a non-parametric manner attempting to combine the strengths of continuous distributions ("random coefficients") and the generality of finite-mixture models (see e.g. McLachlan and Peel, 2004). In a first step, we estimate for each subject the model parameters (preferences α, β, precision λ, and reference point weights w_1, w_2) individually by maximum likelihood.[21] Then, for the predictions that most of our results rely on, we implement a finite mixture approach where each of the n subjects available in-sample has weight $1/n$ out-of-sample. That is, we model the out-of-sample subject pool to be characterized as a finite mixture of n components, each with prior weight $1/n$, where each component corresponds with one subject from the in-sample data set. For illustration, there are 106 subjects in KMR14. The in-sample estimation yields 106 parameter vectors denoted as $(p_1, p_2, \dots)$. This means that the prediction for the other experiments is that with probability $1/106$ a subject has vector p_1, with probability $1/106$ vector p_2 applies, and so on.

The main advantage of this approach that it allows us to capture distributions of parameters and their correlations without parametric assumptions. Any single parameter estimate is somewhat noisy, obviously, but since maximum likelihood estimates are approximately normally distributed, the errors overall cancel out and we obtain a fairly general description of the joint distribution of the individual parameters. The observed reliability of our out-of-sample predictions corroborates this approach. Finally, the approach is equally applicable to all models, also to the models accounting for say warm glow and cold prickle, or envy and guilt, and in this way it allows for an equally general treatment of heterogeneity across models.

Finally, to adjust for the differences in budgets between experiments and the (potential) differences in the weights of round numbers resulting from the differences in options sets, we allow all individual precision parameters λ and the round-number weight κ to be adjusted jointly across subjects when making predictions between experiments. These two scaling parameters are estimated from the data, but this rescaling is applied equally for all models and does therefore not affect the relative ranking. The likelihood-ratio tests of predictive adequacy also follow Schennach and Wilhelm (2016) as described above.

[20] That is, multiples of 100 have focality level $\phi_x = 4$, other multiples of 50 have level 3, other multiples of 10 have level 2, other multiples of 5 have level 1, other integers have level 0, other multiples of 0.5 have level -1 and so on. The results are invariant to positive affine transformations of ϕ, i.e. shifting the level of or scaling ϕ does not affect the results.

[21] For numerical reasons, this step is split up into two substeps. First, we estimate individual preference and precision parameters for all reference point weights satisfying $w_1 \geq w_2$ on a grid of step-size 0.1. Secondly, we determine for each individual the likelihood maximizing reference point weights, taking the "smallest" reference point weights in cases of non-uniqueness (non-uniqueness occurs mainly for subjects consistently maximizing their pecuniary payoffs).

Table 4: Instructions differ in the declaration and strength of assignment of endowments

Experiment	Instructions	Classification
AM02	"[...] you are asked to make a series of choices about how to divide a set of tokens between yourself and one other subject in the room."	neutral
HJ06	"[...] you are asked to make a series of choices about how to divide points between yourself and one other subject in the other room"	neutral
CHST07	"[...] you must decide how you want to divide the joint production between yourself and your opponent. In the example above the contributions of the two players to the joint production are 800 NOK and 200 NOK, respectively."	loaded
KMR12	"The blue player has to decide how much of $Y, a fixed amount of money, to pass to the green player and how much to keep for himself/herself. [...] In addition to the money passed by the blue player, the green player will also earn $X."	loaded
KMR13	"Blue will be asked to make a series of 18 choices about how to divide a set of tokens between herself and the Green player. [...] Each choice that Blue makes is similar to the following: Green has 15 points. Divide 50 tokens: HOLD [blank] @ 1 point(s) each, and PASS [blank] @ 2 point(s) each."	neutral (dictator) loaded (recipient)
List07	"Everyone in Room A and in Room B has been allocated $5. The person in Room A (YOU) has been provisionally allocated an additional $5. Participants in Room B have not been allocated this additional $5.[...] decide what portion, if any, of this $5 to transfer to the person you are paired with in Room B. You can also transfer a negative amount: i.e., you can take up to $1 from the person in Room B."	loaded
Bard08	"Each of you has been given £6. [...] You can either leave payments unchanged, increase your own, by decreasing the other person's payment, or decrease your own, increasing the other person's payment."	loaded
KMR14	"In different scenarios you will decide what portion of your endowment to transfer to another participant in the room. Each scenario specifies how much money is in your endowment, how much money is in the OTHER endowment and the range of allowable transfers. In some scenarios you can also transfer a negative amount: i.e., you can take some of the OTHER endowment."	loaded
LMW12	"You will have to decide how to distribute €10 between yourself and the person."	neutral

C Robustness checks in the econometric analysis

The purpose of this section is to show that the results are highly robust to variations in the three econometric assumptions: functional form for reference points (Assumption 3), relative focality of the numbers that may be entered (Footnote 20), extent of round-number effects ($\kappa = 0.8$ in Eq. (15)).

Result 4 (Summary of the robustness checks)**.**

- *We examine four different specifications clarifying how reference points change across contexts (see Definitions 3–5). In line with the theoretical prediction that welfare-based altruism improves model adequacy for all reference point specifications, both descriptive and predictive adequacy (in-sample and out-of-sample) improve highly significantly for all specifications. See Table 5, panel "Aggregate".*
- *We examine two alternative specifications for factoring out round-number effects, the results are very similar for all specifications as shown. See Tables 6 and 7 in comparison to Table 5.*
- *Throughout, we allow for* non-linear inequity aversion *as third benchmark model to extend payoff-based CES altruism. This extension fits substantially worse than the standard linear one examined above and hence was not reported in the paper. See the lines "+ Inequity Aversion (nonl)" in all the tables referenced above.*

C.1 Definitions

For clarity, we first repeat the (deliberately simplistic) base model from the main text.

Definition 3 (Welfare-based altruism (base model))**.** In game $\Gamma = \langle B_1, B_2, P_1, t\rangle$, using $w_1, w_2 \in [0,1]$,

$$r_1(\Gamma) = w_1 \cdot B_1 + w_2 \cdot tB_2$$
$$r_2(\Gamma) = w_2 \cdot B_1 + w_1 \cdot tB_2.$$

Our second robustness check is a model similar to Definition 3, but other endowments are weighed by transfer rate. This implicitly yields inequity averse reference points for $w_1 = w_2$ (scaled down or up if $w_1 + w_2 \gtrless 1$). It is equivalent to Definition 3 if $t = 1$. By comparing it to Definition 3, we can evaluate if subjects take the transfer rate into account when forming reference points. Notable special cases are CES ($w_1 = w_2 = 0$), and inequity aversion/egalitarian ($w_1 = w_2 = 0.5$), strict libertarian ref points ($w_1 = 1, w_2 = 0$). Obviously, the model allows for a continuum in-between.

Definition 4 (Welfare-based altruism 2 (robustness check I))**.** In game $\Gamma = \langle B_1, B_2, P_1, t\rangle$, using $w_1, w_2 \in [0,1]$,

$$r_1(\Gamma) = w_1 \cdot B_1 + w_2 \cdot B_2$$
$$r_2(\Gamma) = w_2 \cdot tB_1 + w_1 \cdot tB_2.$$

Our second robustness check adapts the base model in Definition 3 by allowing for the background income to equate with the minimal payoff, rather than the outside-laboratory payoff.

Definition 5 (Welfare-based altruism 3 (robustness check II))**.** In game $\Gamma = \langle B_1, B_2, P_1, t\rangle$, using $w_1, w_2 \in [0,1]$,

$$r_1(\Gamma) = \min p_1 + w_1 \cdot (B_1 - \min p_1) + w_2 \cdot (tB_2 - \min p_2)$$
$$r_2(\Gamma) = \min p_2 + w_2 \cdot (B_1 - \min p_1) + w_1 \cdot (tB_2 - \min p_2).$$

Our final robustness check is the arguably most realistic model used in the theoretical analysis, weighing by transfer rate and using the minimal payoff as background income. This model usually fits best. It contains status-quo-based reference points ($w_1 = w_2 = 0$) and strict expectations-based reference points ($w_1 + w_2 = 1$) as the most notable special cases, and by allowing for $w_1 + w_2 \in (0,1)$ all convex combinations are also included.

Definition 6 (Welfare-based altruism 4 (robustness check III))**.** In game $\Gamma = \langle B_1, B_2, P_1, t\rangle$, using $w_1, w_2 \in [0,1]$,

$$\begin{aligned} r_1(\Gamma) &= \min p_1 + w_1 \cdot (B_1 - \min p_1) + w_2 \cdot (B_2 - \min p_2/t), \\ r_2(\Gamma) &= \min p_2 + w_2 \cdot t \cdot (B_1 - \min p_1) + w_1 \cdot (t \cdot B_2 - \min p_2). \end{aligned}$$

As non-linear model of inequity aversion, we use the following straightforward extension of CES altruism.

Definition 7 (Non-linear inequity aversion)**.** Using the notation in the main text, non-linear inequity aversion is defined as follows:

$$u(\pi) = (1 - \alpha_1 - \alpha_2 - \alpha_3) \cdot \pi_1^\beta + \alpha_1\, \pi_2^\beta - \alpha_2 \cdot |\pi_1 - \pi_2|_+^\beta - \alpha_3 \cdot |\pi_2 - \pi_1|_+^\beta. \tag{+ Inequity Aversion (nonl)}$$

Finally, as simplified focality weights as robustness check for the standard focality weights described above (Footnote 20, which follows Breitmoser (2017)), we use the following.

Definition 8 (Simplified focality weights)**.** All numbers that are multiples of 5 have focality weight $\phi = 1$ in Eq. (15), all other numbers have focality weight $\phi = 0$.

C.2 Results

Table 5: Predictions for standard focality weights and $\kappa = 0.8$ (results from main text)

		Descriptive	Predictive	Details on predictions of ...			
Calibrated on	Altruism is ...	Adequacy	Adequacy	Dictator	Endowments	Taking	Sorting
Aggregate	Payoff based (CES)	5839.8	27404.1	9343.1	9631.8	5546.8	2882.4
	+ Warm Glow/Cold Prickle	5354.6^{++}	28075^{--}	9896.5^{--}	9581.6	5617.8^{-}	2979.1^{--}
	+ Inequity Aversion	5453.9^{++}	27447.9	9094^{+}	9859.8^{--}	5600.4^{--}	2893.7
	+ Inequity Aversion (nonl)	5718.2^{+}	27435	9196.1	9811.9^{--}	5546.4	2880.6
	Welfare based	5035.7^{++}	26674.4^{++}	9093.2^{++}	9385^{++}	5451^{++}	2745.2^{++}
	Welfare based (adj)	5035.7^{++}	25740.4^{++}	8883.6^{++}	9023.5^{++}	5212.2^{++}	2631.2^{++}
	Welfare based 2	5181.4^{++}	26919.5^{++}	9108.6^{++}	9529.5	5473.3^{+}	2808.2^{++}
	Welfare based 2 (adj)	5181.4^{++}	26209^{++}	8852.9^{++}	9179.9^{++}	5393.2^{++}	2793^{++}
	Welfare based 3	5048.4^{++}	27064.9^{+}	9221.4	9640.7	5494.5^{+}	2708.2^{++}
	Welfare based 3 (adj)	5048.4^{++}	25920^{++}	8559.7^{++}	9306.4^{++}	5393.2^{++}	2670.7^{++}
	Welfare based 4	4936.9^{++}	26945^{++}	9308.3	9354.1^{++}	5493.6^{+}	2789^{++}
	Welfare based 4 (adj)	4936.9^{++}	25703.9^{++}	8594.5^{++}	9167.1^{++}	5286.7^{++}	2665.6^{++}
Dictator games	Payoff based (CES)	1460.9	8950.5	1343.4	4339	2353.3	914.7
	+ Warm Glow/Cold Prickle	1507.3^{--}	8854.6	1343	4218.4^{+}	2375.2	917.9
	+ Inequity Aversion	1234.6^{++}	8794.8^{++}	1217.1^{+}	4311.7	2360.7	905.3
	+ Inequity Aversion (nonl)	1314.9^{++}	8943.8	1271.4^{++}	4391.2^{--}	2357.8	923.3
	Welfare based	1146.6^{++}	8758^{++}	1279.8^{+}	4273.8^{+}	2316.6^{+}	887.7
	Welfare based (adj)	1146.6^{++}	8603.9^{++}	1263.9^{+}	4152.5^{++}	2300.8^{++}	888.2
	Welfare based 2	1146.4^{++}	8849.2^{+}	1276.4^{+}	4355.8	2325^{+}	892^{+}
	Welfare based 2 (adj)	1146.4^{++}	8585.3^{++}	1265.7^{+}	4119.5^{++}	2309.7^{++}	892^{+}
	Welfare based 3	1055^{++}	8818.4^{+}	1272.6^{+}	4336.7	2321.5^{+}	887.5^{+}
	Welfare based 3 (adj)	1055^{++}	8673.6^{++}	1255.2^{+}	4231.6	2307.8^{+}	880.5^{+}
	Welfare based 4	1050.9^{++}	8715.2^{++}	1268.8^{+}	4240.1^{++}	2324.2	882.1^{++}
	Welfare based 4 (adj)	1050.9^{++}	8662.1^{++}	1252.5^{+}	4219.8^{++}	2309.4^{+}	881.9^{++}
Gen Endowments	Payoff based (CES)	2896.6	8752.9	4260.4	826.1	2613.8	1052.7
	+ Warm Glow/Cold Prickle	2395.5^{++}	8967.8^{--}	4289.6	954.5^{--}	2649.7	1074
	+ Inequity Aversion	2800.1^{+}	8916.4^{--}	4333.6^{-}	849.9	2663^{--}	1069.9^{--}
	+ Inequity Aversion (nonl)	2923.3	8703.6^{+}	4235.2	824.5	2599.8^{+}	1044^{++}
	Welfare based	2662.7^{++}	8416.7^{++}	4084.2^{++}	767.9^{+}	2565.9^{+}	998.7^{++}
	Welfare based (adj)	2662.7^{++}	7867.7^{++}	3985.8^{++}	637.1^{++}	2351^{++}	895.4^{++}
	Welfare based 2	2769.6^{++}	8615.1^{++}	4157.7^{+}	819.4	2580.2	1057.8
	Welfare based 2 (adj)	2769.6^{++}	8312.5^{++}	3995.9^{++}	751.5^{++}	2521.6^{++}	1045.1
	Welfare based 3	2730^{++}	8626.1^{++}	4236.6	822.1	2606.2	961.2^{++}
	Welfare based 3 (adj)	2730^{++}	7928.2^{++}	3692.7^{++}	778.5^{+}	2524.5^{++}	934^{++}
	Welfare based 4	2662.7^{++}	8754.3	4319.1^{-}	782	2601.1	1052
	Welfare based 4 (adj)	2662.7^{++}	7710.2^{++}	3719.7^{++}	643.3^{++}	2413.4^{++}	935.2^{++}
Taking Games	Payoff-based (CES)	1482.4	9700.7	3739.3	4466.7	579.7	914.9
	+ Warm Glow/Cold Prickle	1451.8	10252.5^{--}	4263.8^{--}	4408.7	592.8	987.2^{--}
	+ Inequity Aversion	1419.2^{+}	9736.7	3543.3^{++}	4698.2^{--}	576.6	918.5
	+ Inequity Aversion (nonl)	1479.9	9787.7	3689.5	4596.1^{--}	588.8^{-}	913.3
	Welfare-based	1226.4^{++}	9499.7^{+}	3729.2	4343.2	568.5^{+}	858.8^{++}
	Welfare based (adj)	1226.4^{++}	9270.3^{++}	3633^{+}	4232.9^{++}	559.3^{++}	846.6^{++}
	Welfare-based 2	1265.5^{++}	9455.3^{++}	3674.5	4354.2^{+}	568.2	858.3^{++}
	Welfare based 2 (adj)	1265.5^{++}	9310.1^{++}	3590.4^{+}	4305.4^{++}	560.9^{+}	854.9^{++}
	Welfare-based 3	1263.4^{++}	9620.4	3712.2	4482	566.9	859.4^{++}
	Welfare based 3 (adj)	1263.4^{++}	9312.1^{++}	3603.2^{+}	4295.4^{+}	559.8^{+}	855.2^{++}
	Welfare-based 4	1223.4^{++}	9475.5^{++}	3720.4	4332^{+}	568.2^{+}	855^{++}
	Welfare based 4 (adj)	1223.4^{++}	9331.8^{++}	3620.6	4302.4^{+}	562.9^{++}	847.5^{++}

Table 6: Predictions for simplified focality weights and $\kappa = 0.8$ (robustness check)

		Descriptive	Predictive	Details on predictions of . . .			
Calibrated on	Altruism is . . .	Adequacy	Adequacy	Dictator	Endowments	Taking	Sorting
Aggregate	Payoff based (CES)	5968.4	27868.5	10084.1	9676.9	5277.3	2830.1
	+ Warm Glow/Cold Prickle	5546.9^{++}	28922.2^{--}	10687^{--}	9846.6^{-}	5428^{--}	2960.7^{--}
	+ Inequity Aversion	5593.9^{++}	27994.9	9944	9772.7	5377.8^{--}	2900.4^{--}
	+ Inequity Aversion (nonl)	6128.5^{--}	28905.7^{--}	10752.5^{--}	9827^{--}	5353.9^{--}	2972.4^{--}
	Welfare based	4677.3^{++}	27288.5^{++}	9790.8^{++}	9560.8	5232.4^{+}	2704.5^{++}
	Welfare based (adj)	4677.3^{++}	26308.2^{++}	9618.3^{++}	9107.8^{++}	5000.7^{++}	2591.4^{++}
	Welfare based 2	5023.7^{++}	26894.6^{++}	9920.8^{+}	8986.9^{++}	5275.3	2711.6^{++}
	Welfare based 2 (adj)	5023.7^{++}	26240.1^{++}	9659.8^{++}	8759^{++}	5148^{+}	2683.3^{++}
	Welfare based 3	5258^{++}	27031.7^{++}	9843.9^{++}	9180.6^{++}	5270.6	2736.6^{++}
	Welfare based 3 (adj)	5258^{++}	26174.3^{++}	9591.9^{++}	8875.1^{++}	5133.9^{+}	2583.4^{++}
	Welfare based 4	5258^{++}	26772.1^{++}	9733.1^{++}	9174.7^{++}	5202.5^{+}	2661.8^{++}
	Welfare based 4 (adj)	5258^{++}	25472.9^{++}	8880^{++}	8933.2^{++}	5088.7^{++}	2581^{++}
Dictator games	Payoff based (CES)	1697.2	8998.3	1462.4	4387.3	2253.5	895.1
	+ Warm Glow/Cold Prickle	1715.9	9104.6^{--}	1502.8^{--}	4377.3	2304^{--}	920.4^{-}
	+ Inequity Aversion	1390.2^{++}	8834.1^{+}	1352^{+}	4313.9	2268.5	899.7
	+ Inequity Aversion (nonl)	1753^{--}	9117.8^{--}	1484.9^{-}	4418.9	2295.4^{--}	918.6^{-}
	Welfare based	1392^{++}	8807.9^{++}	1396.7^{++}	4335.2	2208.4^{++}	867.5
	Welfare based (adj)	1392^{++}	8473.5^{++}	1349.4^{++}	4080^{++}	2184.7^{++}	861^{+}
	Welfare based 2	1400.9^{++}	8757.5^{++}	1442.9	4170.8^{++}	2258	885.9
	Welfare based 2 (adj)	1400.9^{++}	8654.1^{++}	1437	4090.9^{++}	2248.6	879.1
	Welfare based 3	1392.3^{++}	8801^{++}	1391.2^{++}	4266^{++}	2270.1	873.7
	Welfare based 3 (adj)	1392.3^{++}	8548.6^{++}	1348.2^{++}	4071.4^{++}	2263.4	867.1^{+}
	Welfare based 4	1392.7^{++}	8707.2^{++}	1360.6^{+}	4234.8^{+}	2257.3	854.4
	Welfare based 4 (adj)	1392.7^{++}	8529.2^{++}	1356.5^{+}	4093.5^{++}	2241.9	838.8^{+}
Gen Endowments	Payoff based (CES)	2870.3	8828.2	4503	840.8	2441.2	1043.2
	+ Warm Glow/Cold Prickle	2438.7^{++}	9018.4^{--}	4518.8	936.9^{--}	2500.4^{--}	1062.4
	+ Inequity Aversion	2837.6	9057.8^{--}	4650.6^{--}	841.7	2504.8^{--}	1060.6^{-}
	+ Inequity Aversion (nonl)	2926.5^{--}	9003.2^{--}	4609.8^{--}	871.4^{--}	2453.2	1068.8^{--}
	Welfare based	2149.8^{++}	8650.6^{++}	4372.5^{++}	836.2	2448.7	993.1^{+}
	Welfare based (adj)	2149.8^{++}	8159.9^{++}	4308.6^{++}	703.3^{++}	2250.8^{++}	898.8^{++}
	Welfare based 2	2387^{++}	8763.2	4561.1	767^{++}	2443.9	991.2^{+}
	Welfare based 2 (adj)	2387^{++}	8321.2^{++}	4347.6^{+}	676.6^{++}	2329.4^{+}	969.1^{++}
	Welfare based 3	2636.8^{++}	8763.8	4544	762.5^{++}	2427.8	1029.4
	Welfare based 3 (adj)	2636.8^{++}	8216^{++}	4362.4^{++}	673^{++}	2299.8^{++}	882.2^{++}
	Welfare based 4	2586.3^{++}	8494.9^{++}	4401^{++}	774.9^{++}	2366^{++}	953^{+}
	Welfare based 4 (adj)	2586.3^{++}	7618.3^{++}	3757.5^{++}	696.4^{++}	2275.2^{++}	890.7^{++}
Taking Games	Payoff-based (CES)	1400.9	10041.9	4118.7	4448.7	582.6	891.9
	+ Warm Glow/Cold Prickle	1392.3	10799.2^{--}	4665.4^{--}	4532.4^{-}	623.5^{--}	977.9^{--}
	+ Inequity Aversion	1366.1^{+}	10103	3941.3^{+}	4617^{--}	604.6^{--}	940.1^{--}
	+ Inequity Aversion (nonl)	1448.9^{--}	10784.7^{--}	4657.8^{--}	4536.7^{--}	605.3^{--}	985^{--}
	Welfare-based	1135.5^{++}	9830^{+}	4021.5	4389.4	575.2	843.9^{+}
	Welfare based (adj)	1135.5^{++}	9676.3^{++}	3959.4^{++}	4323.5^{+}	564.2^{++}	830.7^{++}
	Welfare-based 2	1235.8^{++}	9374^{++}	3916.9^{++}	4049.1^{++}	573.4	834.6^{++}
	Welfare based 2 (adj)	1235.8^{++}	9266.3^{++}	3874.2^{++}	3990.5^{++}	569^{+}	834.1^{++}
	Welfare-based 3	1228.9^{++}	9466.8^{++}	3908.7^{++}	4152.1^{++}	572.7	833.4^{++}
	Welfare based 3 (adj)	1228.9^{++}	9411.2^{++}	3880.3^{++}	4129.7^{++}	569.6^{+}	833^{++}
	Welfare-based 4	1279.1^{++}	9569.9^{++}	3971.4	4164.9^{++}	579.1	854.4
	Welfare based 4 (adj)	1279.1^{++}	9326.9^{++}	3765.1^{++}	4142.3^{++}	570.5	850.5^{+}

Table 7: Predictions for standard focality weights and $\kappa = 0.6$ (robustness check)

		Descriptive	Predictive	Details on predictions of …			
Calibrated on	Altruism is …	Adequacy	Adequacy	Dictator	Endowments	Taking	Sorting
Aggregate	Payoff based (CES)	5858.5	27706.5	9385.7	9895.7	5561.1	2864.1
	+ Warm Glow/Cold Prickle	5385.4^{++}	28483.9^{--}	10056.6^{--}	9809.1	5639^{-}	2979.2^{--}
	+ Inequity Aversion	5458.5^{++}	27412.6^{+}	9048^{+}	9844.8	5636.8^{--}	2882.9
	+ Inequity Aversion (nonl)	5703.4^{++}	27412.1^{+}	9166.5^{+}	9824.5	5548.6	2872.5
	Welfare based	5030.7^{++}	26719^{++}	9156.4^{+}	9359.3^{++}	5480.2^{+}	2723.1^{++}
	Welfare based (adj)	5030.7^{++}	25647.1^{++}	8894.5^{++}	8962.8^{++}	5193.7^{++}	2606.1^{++}
	Welfare based 2	5175.3^{++}	27115^{++}	9350.9	9488.8^{++}	5487.3^{+}	2788^{++}
	Welfare based 2 (adj)	5175.3^{++}	26237.4^{++}	9054^{++}	9037.9^{++}	5402.6^{++}	2752.8^{++}
	Welfare based 3	5015.1^{++}	26985.5^{++}	9207.2	9573.8^{+}	5505.2	2699.3^{++}
	Welfare based 3 (adj)	5015.1^{++}	25725.8^{++}	8509.8^{++}	9191.6^{++}	5401.6^{++}	2632.9^{++}
	Welfare based 4	4927.3^{++}	26759.6^{++}	9189.9	9334.5^{++}	5527	2708.2^{++}
	Welfare based 4 (adj)	4927.3^{++}	25558^{++}	8503.9^{++}	9130.5^{++}	5279.5^{++}	2654.1^{++}
Dictator games	Payoff based (CES)	1493.5	9087.2	1374.2	4442.4	2370.4	900.3
	+ Warm Glow/Cold Prickle	1533	9012.6	1369.2	4355.5^{+}	2378	910
	+ Inequity Aversion	1238.4^{++}	8835.5^{++}	1204.7^{++}	4341.1^{+}	2386.6	903
	+ Inequity Aversion (nonl)	1326.8^{++}	8999.9	1245.1^{++}	4472.8	2366.4	915.6^{-}
	Welfare based	1165.2^{++}	8725.1^{++}	1278^{++}	4256.8^{++}	2317.2^{+}	873.1
	Welfare based (adj)	1165.2^{++}	8486.5^{++}	1235.9^{++}	4077.4^{++}	2302.2^{++}	872.4
	Welfare based 2	1168.9^{++}	8738.7^{++}	1285.8^{++}	4245.4^{++}	2334.2^{+}	873.3
	Welfare based 2 (adj)	1168.9^{++}	8580.8^{++}	1256^{++}	4133.4^{++}	2321.1^{+}	871.8^{+}
	Welfare based 3	1066.6^{++}	8756.4^{++}	1270.3^{+}	4286.8^{+}	2321.5^{+}	877.7^{+}
	Welfare based 3 (adj)	1066.6^{++}	8556.3^{++}	1247.2^{++}	4124.1^{++}	2310^{++}	876.4^{+}
	Welfare based 4	1066.7^{++}	8690.2^{++}	1261.5^{++}	4225.2^{++}	2332.5	870.9^{+}
	Welfare based 4 (adj)	1066.7^{++}	8580.2^{++}	1238.9^{++}	4161.1^{++}	2312.8^{+}	868.9^{+}
Gen Endowments	Payoff based (CES)	2867	8696	4197.9	829.2	2613.8	1055.2
	+ Warm Glow/Cold Prickle	2383.2^{++}	9015.5^{--}	4311.2^{--}	961.3^{--}	2668	1075.1
	+ Inequity Aversion	2791.6^{+}	8899.2^{--}	4291.5^{--}	855.6	2681.2^{--}	1070.9^{-}
	+ Inequity Aversion (nonl)	2892	8677.4	4249.7	786.3^{++}	2595.8	1045.7
	Welfare based	2631.6^{++}	8479.8^{++}	4122.8^{+}	769.8^{+}	2586.6	1000.7^{++}
	Welfare based (adj)	2631.6^{++}	7884.6^{++}	4026^{++}	640.5^{++}	2329.2^{++}	890.3^{++}
	Welfare based 2	2731.8^{++}	8809.8^{--}	4348.1^{--}	813.5	2591	1057.2
	Welfare based 2 (adj)	2731.8^{++}	8468.9^{++}	4154	748.4^{++}	2522.8^{++}	1045
	Welfare based 3	2673.4^{++}	8750.8	4315.3^{--}	848.7	2621.3	965.5^{++}
	Welfare based 3 (adj)	2673.4^{++}	7915.3^{++}	3677^{++}	788.2^{+}	2532.8^{+}	918.8^{++}
	Welfare based 4	2626.2^{++}	8624.4	4242.4	776.7^{+}	2618.6	986.7^{++}
	Welfare based 4 (adj)	2626.2^{++}	7705.4^{++}	3715.2^{++}	647^{++}	2403.3^{++}	941.4^{++}
Taking Games	Payoff-based (CES)	1498.1	9923.3	3813.6	4624.1	576.9	908.6
	+ Warm Glow/Cold Prickle	1469.1	10455.7^{--}	4376.2^{--}	4492.4^{++}	593^{-}	994.1^{--}
	+ Inequity Aversion	1428.5^{+}	9677.9^{++}	3551.8^{++}	4648.1	568.9^{+}	909
	+ Inequity Aversion (nonl)	1484.7	9734.8^{+}	3671.7^{+}	4565.4	586.4^{-}	911.2
	Welfare-based	1234^{++}	9514.1^{++}	3755.7	4332.7^{++}	576.5	849.3^{++}
	Welfare based (adj)	1234^{++}	9277.5^{++}	3631.6^{++}	4243.8^{++}	561.3^{++}	842.4^{++}
	Welfare-based 2	1274.6^{++}	9566.5^{++}	3717	4429.9^{++}	562.2	857.5^{++}
	Welfare based 2 (adj)	1274.6^{++}	9187.6^{++}	3643^{++}	4153.4^{++}	557.7^{+}	835^{++}
	Welfare-based 3	1275.1^{++}	9478.3^{++}	3621.6^{++}	4438.3^{+}	562.3	856.1^{++}
	Welfare based 3 (adj)	1275.1^{++}	9255.1^{++}	3584.5^{++}	4278.2^{++}	557.2^{+}	836.7^{++}
	Welfare-based 4	1234.4^{++}	9445.1^{++}	3686	4332.5^{++}	575.9	850.7^{++}
	Welfare based 4 (adj)	1234.4^{++}	9273.4^{++}	3548.9^{++}	4321.1^{++}	562.2^{++}	842.8^{++}