Moral Costs and Rational Choice:
Theory and Experimental Evidence

James C. Cox[a], John A. List[b], Michael Price[c],
Vjollca Sadiraj[a], and Anya Samek[d]

[a]Georgia State University
[b]University of Chicago and NBER
[c]University of Alabama and NBER
[d]University of Southern California

ABSTRACT

Literature exploring other-regarding behavior reveals interesting phenomena, yet less attention has been given to implications for foundational assumptions within economics. Our study synthesizes the evidence, explaining why recent work challenges rational choice theory as well as its special case, convex preference theory. Guided by this understanding, we advance a theory of choice that exhibits monotonicity with respect to observable reference points. This modification of choice theory establishes consistency with otherwise-anomalous data. We explain how our theory organizes extant data and has applications to strategic games with contractions. We report an experiment designed to test central features of the new theory.

1. INTRODUCTION

One of the most influential bodies of economics research in the past two decades revolves around whether and to what extent people value efficiency, fairness, equity, and reciprocity. Experimental work has provided evidence that such motivations can be important in creating and determining surplus allocations in markets (see, e.g., Fehr et al., 1993; Bandiera et al., 2005; Landry et al., 2010; Cabrales et al., 2010; Hertz and Taubinsky, 2017), with accompanying theoretical models of social preferences providing a framework to rationalize such behaviors (see, e.g., Rabin, 1993; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Andreoni and Miller, 2002; Cox, Friedman and Sadiraj, 2008; Fudenberg and Levine, 2012; Celen et al., 2017; Galperti and Strulovici, 2017).

Within this line of research, pro-social preferences have been measured using a class of experiments taking the form of dictator games, gift exchange games, public goods games, ultimatum games, and trust games. Such games have shown that social preferences touch many areas of economic interactions, and the received literature suggests that observed sharing behaviors are consonant with existing theory. For instance, in a seminal study, Andreoni and Miller (2002) show that in a modified dictator game subjects' choices satisfy the key axiom of revealed preference theory.[1] More recently, Andersen et al. (2011) provide data that reveals demand curves for fairness in an ultimatum game are downward sloping.

The shortage of work challenging basic tenets in the sharing literature contrasts sharply with other areas of behavioral economics, which have lent deep insights into foundational assumptions within economics. For example, for riskless choice, received results reveal that many consumers have preferences defined over changes in consumption, but individual behavior converges to the neoclassical prediction as trading experience intensifies (see, e.g., Kahneman et al., 1990; List, 2004; Engelmann and Hollard, 2010).

Relatedly, for choice that involves risk, several scholars (see, e.g., Harless, 1992; Hey and Orme, 1994) present econometric estimates of indifference curves under risk at the individual level that show neither expected utility theory nor the non-expected utility alternatives do a satisfactory job of organizing behavior. Choi et al. (2007) extend this analysis by developing an experimental protocol that allows the researcher to both test the consistency of choices with the assumption of utility maximization and estimate a two-parameter utility function for each individual. These

---

[1] Fisman et al. (2007) extend this earlier work by developing an experimental framework that allows the researcher to not only test the consistency of choices but also recover individual level preferences for giving. Fisman et al. (2015) explore how preferences for giving are impacted by macroeconomic shocks.

examples are not exhaustive, as there are many other active research inquiries in this spirit, including those exploring intertemporal choice (see, e.g., Laibson, 1997; O'Donoghue and Rabin, 1999, 2001; Frederick et al., 2002), asymmetry and transitivity of preferences (Tversky, 1969; Slovic 1995; Cox and Grether, 1996; List, 2002), and conditional altruism (Dufwenberg and Kirchsteiger, 2004; Cox, Friedman, and Sadiraj, 2008).

Our study follows the spirit of this broader literature by exploring whether basic economic tenets are satisfied in sharing choices as observed in the dictator game. To understand more deeply the factors that motivate sharing, a number of scholars have augmented the standard dictator game by varying the feasible action set (e.g., List, 2007; Bardsley, 2008; Cappelen, et al., 2013; Korenok, et al., 2014). These studies report that dictators change their allocations in interesting ways when presented a chance to take as well as to give to others. For example, in the typical dictator game the experiment is framed such that "giving nothing" is the least generous act, and substantial sums of money are given away (Engel, 2011). Yet, research shows that if subjects are allowed to give or take money from the other player, they give much less to the other player on average.[2]

The first goal of our study is to step back from the burgeoning literature and attempt to synthesize what we have learned from the experimental exercises of List (2007) and others. We explain that the traditional dictator game, wherein more than 60 percent of dictators pass a positive amount of money, is consistent with neoclassical convex preference theory (Hicks, 1946; Samuelson, 1947). Yet, more recent results from this literature (e.g., List, 2007; Bardsley, 2008; Cappelen, et al., 2013) provide evidence that challenges convex preference theory.[3] An even more fundamental challenge, to the foundation of rational choice theory, is provided by data from one of the treatments in Korenok, et al. (2014) that is inconsistent with the Contraction Consistency Axiom, which for singleton choice sets is the necessary and sufficient condition for a choice function to be rationalizable by a complete and transitive ordering (Sen, 1971).[4]

Building upon this discussion, we advance an axiomatic theory of moral reference points that is consistent with otherwise-anomalous data from prior experiments. Our theoretical development follows the approach in Cox and Sadiraj (2010) to extend choice theory to accommodate dictator game data that violates a central tenet of conventional theory – in this case,

---

[2] This sentiment is well reflected by Zhang and Ortmann (2014) who report results from a meta-analysis of dictator games that allow a taking option and find, "…an economically and statistically significant negative effect on giving…"

[3] See also experiments by Grossman and Eckel, 2015, Engel, 2011; Korenok et al., 2013; Korenok et al., 2014; Zhang and Ortmann, 2014.

[4] For singleton choice sets, the Contraction Consistency Axiom states that if x is chosen from feasible set F then it will also be chosen from any contraction of set F that contains x. For set-valued choice functions, rationality is equivalent to Sen's (1971) Properties α and β (see below), where Property α is the Contraction Consistency Axiom.

the Contraction Consistency Axiom. A key component of our theory is the identification of moral reference points that are *a priori* observable features of feasible sets and initial endowments.[5]

We then design an experiment to test the defining property of the new theory – monotonicity in choice with respect to the dimensions that define moral reference points. Results from our experiment provide support for the importance of moral reference points on observed patterns of sharing. In contrast, the data are at odds with the standard model of rational choice and any model that assumes convex preferences. We view our study as fitting in nicely with the "theory speaking to experiment and experiment speaking to theory" research culture that has permeated experimental economics for decades.

The remainder of our paper is structured as follows. Section 2 explores the implications for theory of distinct types of dictator games in previous literature that challenge: (a) *homo economicus* convex preference theory; (b) other-regarding convex preference theory; and (c) general rational choice theory. Section 3 presents new theory motivated by distinct features of type (b) and type (c) dictator games. Section 4 presents the design of our experiment intended to discriminate between the new theory and traditional theory. Section 5 presents our experimental results. Section 6 presents implications of our theory for related experiments in Andreoni and Miller (2002), Korenok, et al. (2014), Krupka and Weber (2013), Lazear, Malmendier, and Weber (2012), and Oxoby and Spraggon (2008). Section 7 explains how our theory can be applied to strategic games with contractions and presents applications to moonlighting and investment games and to carrot/stick, carrot, and stick games. Section 8 concludes.

2. WHAT CAN WE LEARN ABOUT THEORY FROM DICTATOR EXPERIMENTS?

2.1 Experiments in which Behavior is Inconsistent with (Universal) Selfish Preferences

Kahneman et al. (1986) was the first to conduct a dictator game experiment in economics, giving subjects a hypothetical choice of choosing an even split of $20 ($10 each) with an anonymous subject or an uneven split ($18, $2), favoring themselves. Three-quarters of the subjects opted for the equal split. The wheels were set in motion for three decades of research examining sharing and allocation of surplus in the lab and field. One stylized result that has emerged from the large literature is that more than 60 percent of subjects pass a positive amount to their anonymous partners and, on average, give more than 25 percent of the total available (Engle, 2011).

---

[5] Moral cost models have been suggested in previous work (e.g., Levitt and List, 2007; DellaVigna et al., 2012; Kessler and Leider, 2012; Ferraro and Price, 2013; Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2015.

Even though some scholars have argued that such giving patterns violate deeply held economic doctrines, it is important to recall that preference order axioms do not uniquely identify the commodity bundles. In a two-commodity case, for example, my preferences may be defined over my hotdogs and my hamburgers. But the same formal theory of preferences can be applied to two commodities identified as my hamburgers and your hamburgers. Identification of the commodities in a bundle is an interpretation of the theory. In this way convex preference theory, either developed as neoclassical preference theory (Hicks, 1946; Samuelson, 1947) or revealed preference theory (Afriat, 1967; Varian, 1982) can be used for agents who are either self-regarding or other-regarding. As such, the received results of giving in standard dictator games do not represent a rejection of convex preference theory. Rather, they represent a rejection of a joint hypothesis: convex preferences and the assumption that preferences are self-regarding.

2.2 Experiments in which Behavior is Inconsistent with Convex Preference Theory

More recently, List (2007) and Bardsley (2008), amongst others, have used laboratory dictator game experiments to explore how choices are influenced by introducing opportunities for the dictator to take from another subject. This line of work does present a challenge for convex preference theory, as we explain.

Consider, for example, Figure 1, which shows data from List (2007) and Bardsley (2008).[6] Previous discussions of List's data have focused on comparing the 29% of choices of 0 in the Baseline (standard dictator game allowing giving up to $5) treatment with the 65% of the choices of -1 or 0 in the Take 1 treatment (standard dictator game augmented to allow taking $1 from the recipient). An implication of convexity is that these figures should be the same – a pattern that the data clearly refutes.[7] Convex preference theory also implies that the choices that are in the interior of the feasible sets for both the Baseline and Take 1 treatments should be the same. The data are also inconsistent with this prediction of convex preference theory. Data from Bardsley (2008) and from the experiment with a representative sample of Danish adult subjects reported by Cappelen, et al. (2013) are also clearly inconsistent with convex preference theory.

Popular models of social preferences, including inequality aversion (Fehr and Schmidt 1999; Bolton and Ockenfels 2000), quasi-maximin (Charness and Rabin 2002), CES (Andreoni and Miller 2002), and egocentric altruism (Cox and Sadiraj 2007), have the same implication as

---

[6] The data for List (2007) are from the *JPE* online appendix.

[7] The initial endowments are the same in the two treatments so we can discuss implications of convex preference theory for either payoffs or transfers.

traditional convex preference theory for comparisons such as the 29% vs. 65% choices in List's experiment. Therefore these models are also called into question by the dictator game data.

[FIGURE 1 ABOUT HERE: Histograms for List and Bardsley Data]

Convexity, however, is not a necessary condition for choice rationality, so comparisons such as the above for the List and Bardsley data do not allow the researcher to draw conclusions about choice rationality. An illustration of rational choices for non-convex preferences is shown in Figure 2. Let the endowment be at point B and the feasible set be $[A,B]$ in the give game and the choice be at point y. In the give or take game, let the endowment be at point B and the feasible set be $[A,C]$ and the choice be at point x.

[FIGURE 2 ABOUT HERE: Example of Choice with non-Convex Preferences]

Choice implied by convex preferences and rational choice are different concepts. Convex preference theory is a special case of rational choice theory that imposes far stronger restrictions on observable choices. Rational choice theory requires that choices satisfy consistency (contraction and expansion) axioms (Samuelson, 1938; Chernoff, 1954; Arrow, 1959; Sen, 1971, 1986). If we let $C(S)$ denote the choice set when the feasible set is S and $C(T)$, be the choice set when the feasible set is T then the Contraction Consistency Axiom (also known as Property α from Sen 1971, 1993) states: For any feasible sets S and T and choice sets $C(S)$ and $C(T)$,

CCA: $[x \in C(S) \text{ and } x \in T \subseteq S] \Rightarrow x \in C(T)$.

In words, any allocation $x \in C(S)$ that is chosen from S is also chosen from any subset T of S that contains x. With single-valued choice functions, CCA is the necessary and sufficient condition for existence of a complete and transitive ordering of choices (Sen, 1971).

The feasible set for the Baseline treatment in List (2007) is a contraction of the set for the Take 1 treatment. Therefore, by CCA, anyone choosing an amount from $0 to $5 in the Take 1 treatment should make the same choice in the Baseline treatment. In contrast to the special case of convex preferences, rational choice theory offers no suggestions for the Baseline treatment if one is observed to choose -$1 in the Take 1 treatment. Rational choice theory: (a) *can* accommodate someone who takes in the Take 1 treatment and gives in the Baseline treatment; but (b) *cannot* accommodate someone who gives different amounts in the Take 1 and Baseline treatments.

The above properties of rational choice theory imply that each of the bars portraying fractions of choices of $0 to $5 in the Take 1 treatment should not be higher than the corresponding bar for choices in the Baseline. With the exception of the bar at $1.50 (corresponding to two observations in the Take 1 treatment), the List (2007) data are consistent with rational choice

theory. Similarly, data shown in Figure 1 from Experiment 2 conducted by Bardsley (2008) are inconsistent with convex preferences but are mostly consistent with rational choice theory; the bar at $1.50 (2 observations) is the only inconsistency with rational choice theory in Experiment 2.

2.3 Experiment in which Behavior is Inconsistent with Rational Choice Theory

Korenok et al. (2014) report a dictator game experiment that explores the effects of changing endowments and varying give and take actions while holding constant the feasible set of payoffs. Figure 3 illustrates five different scenarios in the Korenok et al. experiment. In all five scenarios, the feasible set is the same set of discrete points on the budget line shown in Figure 3. What varies across scenarios is the initial (endowed) allocation of $20 between the dictator and the recipient. We represent these scenarios using the numbered points on the budget line in Figure 3. For example, in scenario 1, the dictator is endowed with $20 and the recipient with $0. In scenario 9, the recipient is endowed with $20 and the dictator with $0. Other endowments used in the experiment are shown at points 3, 6, and 8 on the budget line.

[FIGURE 3 ABOUT HERE: Endowments and Choices in Korenok, et al.]

CCA implies that choices will be invariant to changes in the endowments in the experiment: for any two endowments, the choice sets *F* and *G* are the same set. Let S1 ($4.05) denote the average payoff of $4.05 to the recipient in scenario 1. Using this same convention to reflect payoffs in the remaining scenarios, we have that the average recipient payoffs for the five scenarios are: S1 ($4.05), S3 ($5.01), S6 ($5.61), S8 ($6.59), and S9 ($6.31). The differences between these payoff figures are statistically significant except for the comparison of S8 with S9. The fact that average payoffs differ across endowments is inconsistent with predictions from CCA. Actually, the inconsistency with rational choice theory is even more fundamental that inconsistency with CCA. Choice variability with endowments, when the feasible set is invariant, is inconsistent with existence of a (single-valued) choice function.

3. MORAL MONOTONICITY THEORY

The empirical failure of standard theory with data from these simple dictator games suggests that new theory that formalizes somewhat different empirical content is needed. A framework that has been used to describe giving, taking, and related behaviors builds upon the notion of moral cost (Levitt and List, 2007; List, 2007; Lazear et al., 2012; DellaVigna et al., 2012) or concern for norm compliance (Kessler and Leider, 2012; Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2015). Using this framework, individuals are said to share with others to avoid experiencing moral cost from failing to do so or from taking actions that are deemed socially inappropriate. We put

this approach on an axiomatic foundation that follows from initial work by Cox and Sadiraj (2010).

There are two central features of this approach: (1) definition of Moral Monotonicity Axiom (MMA) that is equivalent to the traditional Contraction Consistency Axiom (CCA) *when* contractions preserve the moral reference point; and (2) definition of moral reference points that are *observable features* of feasible sets. We first define and explain MMA. Subsequently, we develop a concept of moral reference points suggested by features of dictator games that produce data anomalous for traditional rational choice theory.

3.1 Moral Monotonicity Axiom

It is natural to expect that choices are monotonic on moral reference points; that is, the more favorable the moral reference point to an agent the larger the allocation to that agent chosen by himself or another, everything else equal. Let f^* be chosen from some feasible set F and G a subset of F that contains it. Let r^G and r^F denote moral reference points for feasible sets G and F, with choice sets G^* and F^*, and suppose they differ from each other only with respect to the value of dimension i. If the moral reference point in G is more favorable to individual i, then we postulate that no choice from G allocates i less than f^*. Similarly, if the moral reference point in G is less favorable to individual i, then no choice from G allocates i more than f^*. Formalizing this, if we let $\langle\rangle$ denote "not smaller" or "not larger", then for every individual $i (= 1, \cdots, n)$:

> **MORAL MONOTONICITY AXIOM (MMA):**
>
> If $G \subseteq F$, $r_i^G \langle\rangle r_i^F$ and $r_{-i}^G = r_{-i}^F$, then $f^* \in F^* \cap G \Rightarrow g_i^* \langle\rangle f_i^*, \forall g^* \in G^*$

What are the implications of MMA for contractions that preserve moral reference points and contain choices from the bigger set? For such subsets MMA implies that the choice set is a singleton and that conventional axioms of rationality (Sen's 1971 Properties α and β) are satisfied. The modified form of Sen's Property α (a.k.a. CCA) for sets that preserve the moral reference point is[8]

> **PROPERTY α_M**: if $G \subseteq F$ and $r^G = r^F$ then $F^* \cap G \subseteq G^*$

For singleton choice sets, this requires f^* to be *the* chosen allocation in any subset, G of F that contains f^*. Implications of MMA for choices is stated in the following proposition.[9]

---

[8] For non-singleton choice sets, the analogue of Sen's (1971) Property β is Property β_M: if $G \subseteq F$ and $r^G = r^F$ then $G^* \cap F^* \neq \phi$ implies $G^* \subseteq F^*$.

[9] The proof of Proposition 1 in Appendix A also shows that MMA implies Property β_M.

PROPOSITION 1: MMA implies Property α_M

PROOF. See Appendix A.

Thus, for opportunity sets that preserve moral reference points, MMA suffices for choices to be rationalizable. Implications of MMA for a variety of dictator games and for play in strategic games with contractions are discussed in sections 6 and 7.

3.2 Moral Reference Points

Ideas about what may constitute a moral reference point are suggested by the idiosyncratic features of designs of (a) the Korenock (2014) experiment; and (b) the List (2007) and Bardsley (2008) experiments. All of these experiments include taking as well as giving opportunities. The Korenock, et al. experiment varies the dictator's endowment while holding constant the minimum (resp. maximum) payoff of each agent at 0 (resp. 20). Their data suggests that dependence of choices on dictator's endowment in a way not captured by traditional theory is empirically significant.[10] In contrast, the paired baseline and take treatments in each of the List and Bardsley experiments hold constant the dictator's endowment while varying the minimum and maximum payoffs. Taken together, these experiments inform that choice behavior is dependent on (a) the dictator's endowment and (b) the maximum and minimum payoffs available in the game. We define moral reference points in a way that is suggested by this experimental literature.

Our definition of moral reference point incorporates two intuitions into theory of choice: that my moral constraints on interacting with you in "the game" we are playing may depend on (a) my endowed (or initial) payoff in the game and (b) the payoff each of us can receive when the other's payoff is maximized (a.k.a. our "minimal expectation payoffs"). Intuition (a) reflects the idea that my moral cost from making a choice that benefits me at your expense decreases with the closeness of my final payoff to my endowed ("status quo") payoff: my "property right."[11] Intuition (b) reflects the idea that my moral cost from making such a choice:

---

[10] In the various treatments, the *sum* of the dictator's and recipient's endowments is held constant. Hence the dependence on endowment could instead be defined on dictator's endowment. Assumed dependence on *both* endowments would be an over-determined miss-specification because they sum to a constant.

[11] The intuition that "property rights" matter for final allocations in a dictator game is consistent with results in Oxoby and Spraggon (2008) and Korenok et al. (2017) who show that dictators share more with the recipient when the total endowment was earned by the actions of the recipient as opposed to the dictator themselves. It is also consistent with results from the meta-analysis in Engel (2011) showing that transfers in dictator games are significantly greater when the recipient earns the endowment.

(i)   *decreases* with the (positive) difference between your final payoff and your minimal expectation payoff − how much more do I give you than the minimum you can expect from the game; and

(ii)   *increases* with the (positive) difference between my final payoff and my minimal expectation payoff − how much more do I give myself than the minimum I can expect from the game.

We now formalize these intuitions and present a concept of moral reference points that are determined by *observable* features of feasible sets. For simplicity, we first use dictator games to illustrate concepts but the model has more general applicability, as explained in section 7 on strategic games with contractions. Our many applications of theory in this paper will all be to two-agent (dictator and strategic) games, but the definition of moral reference point can be extended to n-agent environments, as shown in appendix B.

Let (m, y) denote an ordered pair of payoffs in which my payoff, m is that of the dictator and your payoff, y is that of the recipient. Let the dictator's opportunity set be a finite set F.

Let m^o and y^o be the maximum feasible payoffs for the dictator and the recipient, that is

$$m^o = \max\{m : (m, y) \in F\} \text{ and } y^o = \max\{y : (m, y) \in F\}$$

The minimal expectations point, (m_*, y_*) is defined by the dictator's minimum payoff when the recipient gets their maximum feasible payoff, y^o and the recipient's minimum payoff when the dictator gets their maximum feasible payoff m^o, i.e.

$$m_*(F) = \min\{m : (m, y^o) \in F\} \text{ and } y_*(F) = \min\{y : (m^o, y) \in F\}.$$

Moral cost depends on the minimal expectations point as well as payoff entitlement from the decision maker's endowment. We propose as a moral reference point an ordered pair that agrees with the minimal expectation on the second (recipient's) payoff dimension and is a convex combination of the minimal expectation and the initial endowment e_m on the first (dictator's) payoff dimension. Formally,

(∗) $r^F = ((1-\theta)m_*(F) + \theta e_m, y_*(F)),$

for some $\theta \in (0,1)$. The weight on initial endowment may depend on a variety of things (such as whether endowments were earned or assigned) but all of the analysis in this paper holds for any value of $\theta \in (0,1)$. We use $\theta = 1/2$ in examples only because this makes it easy to visualize results.

An illustration on how to locate moral reference points is provided here for the Give, Take, and Symmetric action sets and Equal, Inequality, and Envy endowment treatments shown in Figure 4. With such downward-sloping budget lines, a moral reference point can be located by: (a) first, find the minimal expectations point, (m_*, y_*) by constructing a right triangle with the budget line as the hypotenuse and the vertical and horizontal sides below and to the left of the budget line; (b) second, find the midpoint of the line segment joining (m_*, y_*) and e (the endowment), and (c) finally, orthogonally project the midpoint onto the line segment joining (m_*, y_*) and the most selfish point. We illustrate this algorithm in Figure 4 for the Equal treatment in the experimental design explained in full in section 4.

[FIGURE 4 ABOUT HERE: Moral Reference Points (Equal Treatments)]

In the Equal-Symmetric treatment the endowment is at B_Q and the feasible set contains discrete points on the budget line extending from A_Q to C_Q. The minimal expectations point is the lower left corner of the large triangle and the moral reference point is r_S. In the Equal-Give treatment the endowment is at point C_Q and the budget line extends from B_Q to C_Q. The minimal expectations point is the lower left corner of the small triangle and the moral reference point is r_G. Finally, in the Equal-Take treatment the budget line extends from B_Q to C_Q. In this case the endowment is at B_Q, so the minimal expectation and moral reference point are both at r_T.

3.3 MMA and Data from Dictator Games in the Literature

As noted above, data from the Korenock, et al. (2014) experiment are inconsistent with existence of any choice function, in particular with rational ones characterized by CCA. In contrast, their data are consistent with MMA. In all of their treatments, the minimum expectations point is the natural origin (because the fixed budget line intersects both axes). Therefore, changes in moral reference points in their design are entirely determined by changes in endowment. The moral reference points defined as in (*) for their several endowment treatments are shown in Figure 3 using $\theta = \frac{1}{2}$. As the endowments move northwest along the budget line the moral reference points move westwards along the horizontal axis from r_1 to r_3 to r_6 to r_8 to r_9, favoring the dictator less and less. Moral monotonicity axiom, MMA requires dictator's choices to decrease the amount allocated to oneself from scenario 1 to 9, a pattern observed in this experiment.

Turning attention back to the experiments reported by List (2007) and Bardsley (2008), we note that while their data are consistent with CCA they also are consistent with MMA. Their experimental designs, however, have little power for testing either CCA or MMA. We next explain the design of an experiment intended to test MMA.

4. EXPERIMENTAL DESIGN AND PROTOCOL

4.1 Experimental Design

We now explain the design of an experiment that provides a direct test of the empirical implications of MMA. Following List (2007), our design begins by introducing an action set in which the dictator can either give to or take from the recipient's initial endowment and compares outcomes in this augmented game to those observed in dictator games in which the participant can only give to, or take from, the recipient. We extend this line of inquiry by considering treatments that vary the initial endowments but preserve the permissible set of actions.[12] If choices are motivated by final allocations only, as assumed in conventional rational choice theory, variation in the initial endowments within a given feasible set should have no impact on observed dictator behavior.

Figure 5 shows three budget lines labeled "Inequality," "Equal," and "Envy." The finite feasible sets include discrete points on the lines. Labelling of the feasible sets reflects the location of the midpoints B_j, $j = I, Q, E$, on the lines. The Symmetric treatments have endowment at B_j and permit the dictator to give (move the allocation towards A_j) or take (move the allocation towards C_j). The Take treatments have endowment at B_j and permit the dictator to take (move the allocation towards C_j). The Give treatments have endowment at C_j and permit the dictator to give (move the allocation towards B_j). There are two prominent features of this design: (a) the corresponding Take and Give treatments have the same feasible set $[B_j, C_j]$; and (b) a Symmetric treatment's feasible set $[A_j, C_j]$ contains the corresponding Take and Give feasible set $[B_j, C_j]$ as a proper subset (a strict contraction).

[FIGURE 5 ABOUT HERE: Feasible Sets]

The experimental design is 3×3: (Inequality, Equal, Envy) × (Symmetric, Take, Give). In the Inequality-Give treatment (with endowment at point C_I in the left panel of Figure 5): the

---

[12] These treatments build upon work by Korenok et al. (2014) and Grossman and Eckel (2015), who employ a variant of the dictator game to explore the effect of give or take actions on choices.

recipient has an endowment of 3; the dictator has an endowment of 27 and can give up to 8 to the recipient. In the Inequality-Take treatment (with endowment at point B_I in the left panel): the recipient has an endowment of 11; the dictator has an endowment of 19 and can take up to 8 from the recipient. In the Inequality-Symmetric treatment (with endowment at point B_I in the left panel): the recipient has an endowment of 11; the dictator has an endowment of 19 and can give up to 8 or take up to 8. The Equal and Envy treatments change the locations of the (point B or point C) endowments but preserve the Give, Take, or Symmetric action sets. In the Equal feasible set, the Symmetric and Take endowment (at point B_Q in the middle panel) is 15 for the recipient and 15 for the dictator. In the Envy feasible set, the Symmetric and Take endowment (at point B_E in the right panel) is 19 for the recipient and 11 for the dictator.[13]

In the Inequality-Symmetric and Envy-Give treatments, the dictator faces an allocation decision over a budget line that crosses the 45-degree line, as in most standard dictator games. In the Equal-Take and Equal-Symmetric treatments, the initial endowment lies on the 45-degree line. However, the treatments differ in that the budget line for the Equal-Take treatment lies on and below the 45-degree line whereas the budget line for the Equal-Symmetric treatment crosses the 45-degree line.

The nine treatments provide a test of the central properties of our theory: monotonicity of choice in both dimensions of moral reference point. Figure 6 shows the moral reference points for our treatments. Note how they vary across treatments along horizontal and vertical lines.

[FIGURE 6 ABOUT HERE: Treatment Moral Reference Points]

4.2 Protocol

The experiment was conducted in the laboratory of the Experimental Economics Center at Georgia State University using students recruited from the student body at Georgia State. When they agreed to participate, subjects knew only that they would be in an economics experiment, but not the exact nature of the experiment. Subjects were given as much time as they wanted to read instructions on their computer monitors. After they were finished reading, summary instructions were projected on a screen and read aloud by an experimenter to make clear that all subjects were given the same information about the decision task. All subjects participated in two practice dictator decisions

---

[13] Note that the sum of the dictator's and recipient's endowments is $30 in all treatments. Thus, as noted above for the List (2007), Bardsley (2008), and Korenok et al. (2013) experiments, it would make no sense to assume dictator's choices are dependent on *both* dictator's and recipient's endowments.

without payoffs to become familiar with both the underlying allocation task and the computer interface. No information was given to subjects about others' practice decisions. After the practice decisions were completed, subjects were informed that the computer would randomly assign them to be active decision makers or passive recipients and that this information would appear on their screen before the start of the first actual round of play.

Subjects were further informed that each active subject would make two decisions while paired with the same recipient and that one of the two decisions would be randomly selected for payoff once both decision rounds were completed. It was explained that these pairings were anonymous and that participants would not know the identity of the person with whom they were paired. A subject made decisions in Give and Take action sets for the same (Equal or Inequality or Envy) setting; or the subject made decisions in Symmetric and Give or Take action sets for the same setting. The order of the games each active subject faced was independently randomly selected. Subjects were asked to complete a short survey after all decisions were made. Once all subjects had completed the survey, they were paid individually and in private their earnings for the chosen decision round. Subject instructions and the survey are available online: http://excen.gsu.edu/jccox/instructions.

5. EXPERIMENTAL RESULTS

5.1 Overview

612 subjects (306 dictators) participated in the experiment. None of the dictators had previous experience (as either dictator or recipient) in dictator games. Each session lasted approximately 50 minutes and each dictator made two decisions. The actual payoffs (from the randomly selected payoff rounds) for dictators were: $19.46 (average) with the range $8 (minimum) to $27 (maximum). Average payoffs and transfers[14] for all data from nine treatment cells are reported in Table 1. Average transfers varied across treatments from a low $2.06 to three times as much, $6.12. Average recipient payoffs varied across treatments from $7.19 to about twice as much, $13.64.

Less than 1/3 (166 out of 612) of observed choices correspond to the most selfish feasible options and less than 1/5 (57 out of 306) of dictators appear selfish (i.e., choose the most selfish option in both decisions). Data exhibit egocentric altruism (Cox and Sadiraj, 2007) as almost all

---

[14] "Transfer" is defined as the amount by which the recipient's payoff exceeds her minimal expectations payoff. In a Give treatment, the transfer is the amount the dictator gives to the recipient. In a Take treatment, the transfer is the amount the dictator does *not* take from the recipient. In a Symmetric treatment, the transfer is the amount not taken plus the amount given, if any.

choices (98%) are such that the dictator's final payoff (weakly) exceeded recipient's final payoff.[15] All data from Give and Take treatments with feasible sets $[B_k, C_k]$ are usable for testing conventional rational choice theory and modified rational choice theory incorporating the Moral Monotonicity Axiom (MMA). Choices from $[A_k, B_k)$ in Symmetric treatments are not usable for testing rational choice theory but there are few such choices; 94% (575 out of 612) of choices are from sets $[B_i, C_i]$ and are thus usable in testing CCA and MMA.

We begin with tests of conventional rational choice theory and MMA using only within-subjects choice pairs. Subsequently, we conduct across-subjects data analysis to ascertain whether observed dictators' transfers are affected by moral reference points as predicted by MMA.

5.2 Consistency of Dictators' Choices with Theoretical Models.

Each dictator made two decisions from the same (Inequality, Equal or Envy) environment: 96 dictators faced budget set $[B_k, C_k]$ twice, in one Give action set and one Take action set; a different group of 98 dictators faced budget set $[A_k, C_k]$ in the Symmetric action set and budget set $[B_k, C_k]$ in the Give action set; and another group of 112 dictators faced budget set $[A_k, C_k]$ in the Symmetric action set and budget set $[B_k, C_k]$ in the Take action set. We created a new variable, Consistency, that takes value 0 if individual's two decisions agreed with theoretical predictions. When the theoretical prediction follows from CCA, 50.19% (135 out of 269; 95% C.I. is [0.44, 0.56]) of subjects made decisions that are consistent with CCA.[16] When the theoretical prediction follows from MMA, 78.07% (210 out of 269) of subjects made decisions that are consistent with MMA. The 95% C.I. of fraction of choices consistent with CCA and MMA are, respectively, [0.44, 0.56] and [0.73, 0.83]. We conclude that:[17]

RESULT 1: MMA organizes our data better than CCA

5.3 Testing MMA: Changes in the Recipient's Moral Reference Dimension

A different test of our theory uses between-subjects data. MMA predicts that the recipient's payoff increases in r_2 when r_1 is fixed. In contrast, convex preferences or CCA predict that changes in

---

[15] Dictator's final payoff was strictly larger than the recipient's payoff in 80% (489 out of 612) of choices.
[16] Data from 37 subjects who gave in the Symmetric treatments are not included as CCA (or MMA) makes no predictions for their choices in Give/Take treatments.
[17] An earlier, working paper version of this paper (Cox et al. 2016) reports data for child subjects in a similar experimental design to test the importance of moral reference points on choices. As with student data reported here, data for children show that final allocations depend on both initial endowments and feasible actions. As such, dictator choices by children violate the standard model of rational choice and any model that assumes convex preferences but provide support for MMA.

r_2 should have no effect on the recipient's payoffs. To test for r_2 effects, we need to look across environments and subjects.

To evaluate whether the recipient's minimal expectations point influences allocations as predicted under MMA, we estimate two Tobit models – one that conditions choice solely upon r_2 and a second that augments this model to include demographic controls for the dictator (gender, race, GPA, religion, major, study year). Each model controls for potential budget constraints (common support across games with a given r_1) by setting as a lower bound the lowest possible payoff a recipient could receive in the common support and as an upper bound the highest possible payoff a recipient could receive in the common support.

For standard models, the estimated coefficient on r_2 should be equal to zero whereas MMA predicts that recipient payoffs are increasing in r_2 and thus a positive coefficient on this treatment parameter. Table 2 presents results for the subset of choices from budget sets with $r_1 = 15$.

[TABLE 2 ABOUT HERE: Tests for Effects of Recipient Moral Reference Dimension]

There are three treatments with the same $r_1=15$ but three different r_2 levels: Inequality-Symmetric ($r_2=3$), Equal-Take ($r_2=7$) and Envy-Give ($r_2=11$), as shown in Figure 6. The recipient's average payoffs across the three treatments (see Table 1) increase as r_2 increases: 9.12 (Inequality-Symmetric), 10.17 (Equal-Take) and 13.43 (Envy-Give). The feasible payoffs for the recipient in these three treatments are integers in the sets: [3, 19] in Inequality-Symmetric, [7, 15] in Equal-Take and [11, 19] in Envy-Give. The budget sets for Envy-Give and Equal-Take are both contractions of the Inequality-Symmetric budget set. Note that set [11, 15] is included in all three treatments. To control for constraints of budget sets on choices, we run Tobit regressions of recipients' final payoffs on data from the three treatments using 11 as the lower bound and 15 as the upper bound. Table 3 reports Tobit estimates of the coefficient on r_2 using models with and without demographic control variables. Consistent with MMA, the estimates for r_2 are positive (p<0.001), which rejects the null hypothesis from CCA (that the estimate is 0) in favor of the alternative hypothesis from MMA. Tests using between-subjects data for $r_1 = 11$ and $r_1 = 19$ lead

to similar conclusions (using data from the two treatments with each of these r_1 values).[18] Thus, using between-subjects data we conclude that:

> **RESULT 2:** The experimental data are consistent with monotonicity in recipient's moral reference dimension.

5.4 Testing MMA: Moral Reference Points and Transfers

We next explore the effects of changing moral reference points on transfers, which capture the dictator's choices as defined in terms of "giving."[19] Differences in the support of feasible budget sets across environments confounds our ability to use payoffs to test implications of MMA.[20] To see the problem, note for example that the dictator's payoff that corresponds to the most selfish feasible choice decreases from \$27 (Inequality) to \$23 (Equal) and down to \$19 (Envy). Looking at transfers (rather than payoffs) makes all data comparable because the set of feasible transfers is invariant across our three environments (Inequality, Equality and Envy).

The feasible set of transfers is [0,8] in both Give and Take action sets and [0,16] in the Symmetric action set for all three environments. Appendix C provides detailed derivations of the implications of MMA and conventional theory for the effects of changing r_1 and r_2 on transfers. However, the basic intuition underlying these formal derivations is as follows. In terms of (recipient moral reference point dimension) r_2 and dictator's payoff m, any feasible transfer, t satisfies $t = 30 - r_2 - m$.[21] Conventional theory (CCA or convex preferences) requires that the dictator's chosen allocation, (m^*, y^*) be preserved in all budget sets *that contain it*. Because the sum of payoffs is constant, such preservation is equivalent to keeping m^* constant, and therefore, under CCA, we expect (a) zero r_1 effect, and (b) negative one-to-one r_2 effect on choice of transfer amount, t^*. In contrast, as shown in appendix C, MMA predicts: (c) negative r_1 effect, and (d) negative smaller (between -1 and 0) r_2 effect on choice of transfer amount, t^*.

---

[18] Tobit regression of recipient's payoff with demographics included in the list of regressors; estimated coefficients of the dummy on the larger value of r_2 are: 1.31 (p-value=0.032) for $r_1 = 11$ and 1.56 (p-value=0.080) for $r_1 = 19$.

[19] That is, in the Give action sets the transfer is the recorded subject's choice. In the Take action set, taking x is by design equivalent in terms of recipient's payoff to "giving" 8-x, hence the transfer is 8 - x. Similarly, in the Symmetric action set, the subject's choice in terms of "giving" is 8-x if the subject takes x and 8+z if the subject gives z.

[20] By design, feasible budget sets shift northwest to the advantage of the recipient as we move from Inequality to Equal and then from Equal to Envy.

[21] The recipient's final payoff is $y = r_2 + t$ in every treatment and the dictator's payoff is $m = 30 - y = 30 - r_2 - t$.

We have the following testable hypotheses that allow us to evaluate whether transfers in our experiment are better organized by CCA or MMA:

H_{r_1} : Marginal effects of r_1 on transfers: 0 (CCA) or negative (MMA)

H_{r_2} : Marginal effects of r_2 on transfers: -1 (CCA) or between -1 and 0 (MMA)

The mean transfers are 4.99 (Inequality, r_2=3), 3.24 (Equal, r_2=7) and 2.37 (Envy, r_2=11).[22] This decreasing pattern is predicted by both CCA and MMA. However, the rate of decrease is one-half of the size predicted by CCA. Table 2 reports results of a Tobit regression that allows us to estimate the effect of changing moral reference points on observed transfers. The list of regressors includes dictator's (r_1) and recipient's (r_2) coordinates of moral reference points of budget sets and, in model (2), demographic controls. As each dictator made two choices, we cluster standard errors. As a robustness check, Table 2 also presents results from a Hurdle model (Cragg, 1971) which allows for the effects of moral reference points to differ along the extensive (whether to make a positive transfer) and intensive (the amount of any positive transfer) margins.

As noted in the first row of Table 3, the estimated coefficient on r_1 is negative and significantly different from 0. The dependence of transfers on r_1 rejects CCA in favor of MMA. The estimate of r_2, in row two of Table 3, is negative. The Wald test rejects the CCA hypothesis that the estimate equals -1.[23] The estimates are consistent with MMA and are robust to both the inclusion of demographics in the regression and the use of a Hurdle model.

[TABLE 3 ABOUT HERE: Testing r_1 and r_2 Effects on Transfers]

This provides our second result on effects of moral reference point on transfers:

RESULT 3: CCA is rejected in favor of MMA.

In summary, our data provides empirical support for MMA predictions of how changes in the moral reference points affect transfers.

5.5 Alternative Models

We briefly look at implications of alternative models of behavior: selfish or social preferences, reference dependence (Koszegi and Rabin 2006), and sharing and sorting (Lazear, et al. 2012).

---

[22] Kruskal-Wallis test: chi-squared=30.25, p-value=0.001; for each subject, the data point is the mean of two transfers.
[23] $F(1,610)$=30.50 (p-value=0.000) for the test for effect of change in $r_1 = 0$. For the joint CCA hypothesis (effect of change in $r_1 = 0$ and effect of change in $r_2 = -1$), $F(2,610)$=39.03 (p-value=0.000).

Selfish Preferences: Two-thirds of the transfers are positive and four out of five of our dictators made at least one positive transfer. Any feasible non-zero transfer reduces a dictator's payoff; so, this model predicts a null transfer, and hence, changes in r_1 or r_2 will have no effect. Parameter estimates for both r_1 and r_2 are statistically significant, rejecting selfish behavior.

Convex Social Preferences: All prominent models of social preferences, including inequality aversion (Fehr and Schmidt 1999; Bolton and Ockenfels 2000), quasi-maximin (Charness and Rabin 2002), CES (Andreoni and Miller 2002), and egocentric altruism (Cox and Sadiraj 2007) assume convex upper contour sets. Our data reject convex preference theory, so these social preferences models are also rejected.

Reference Dependent Model. Koszegi and Rabin (2006) develop a model of reference dependence that has recently seen a surge in applied work. Predictions of this model for our games are similar to standard rational choice theory because, in deterministic settings, optimal "consumption" derived for the conventional preferences model is the "preferred personal equilibrium" in the reference dependent model.[24] Because our data reject conventional theory, the reference dependent model is also rejected.

Sharing and Sorting. Lazear et al. (2012) offer a model of sharing that depends on the environment, $u(D, m, y)$ where D takes value 1 when the environment allows sorting and 0 otherwise. In all of our treatments sorting is not available (i.e., people cannot sort in or out of participating in the games), hence implications of their model for play in our games are similar to standard preference theory, which is rejected by our data. These comparisons suggest:

RESULT 4: The experimental data is inconsistent with an array of behavioral models that have been used to explain sharing behavior.

To summarize, our data provide evidence at odds with standard rational choice theory; the data are also at odds with a suite of alternative behavioral models that have been used to explain sharing. Viewed in its totality, we thus believe our data provides compelling evidence that objectively defined moral reference points matter and influence choice in a manner consistent with MMA.

6. IMPLICATIONS OF MMA FOR OTHER TYPES OF DICTATOR GAMES

To formalize the ways in which moral reference points may influence decision-making in dictator games, we introduced the Moral Monotonicity Axiom (MMA) and applied it to analyze data from our experiment. Yet, MMA has broader implications for choice in a range of related experiments

---

[24] See Proposition 3 in Koszegi and Rabin (2006, pg.1145).

including standard (give-only) dictator games (Andreoni and Miller 2002), other dictator games that compare the effect of give versus take actions on choices (Korenok et al. 2014; Cox et al., 2016), the "bully" dictator game (Krupka and Weber 2013), dictator games with outside options (Lazear, Malmendier, and Weber 2012), and dictator games where property rights and endowments are earned (Oxoby and Spraggon, 2008; Korenok et al., 2017).

6.1 MMA and WARP

As previously mentioned, Andreoni and Miller (2002) conducted dictator game experiments that varied underlying budget sets and applied the generalized axiom of revealed preference (GARP) to analyze the consistency of choices in their setting. Figure 7 illustrates two budget sets like those that the dictator can face in the Andreoni and Miller design. Let point a denote the endowment on the steeper line and point b denote the endowment on the flatter line. Further, consider the shaded quadrilateral that is the intersection of sets bounded by the steeper and flatter budget lines. Viewed through the lens of MMA, the shaded quadrilateral set can be considered a feasible set with endowment at point a. The minimal expectations point is the origin (0,0) for all three feasible sets. Therefore, the moral reference points for the three feasible sets are on the horizontal axis, halfway between 0 and the respective endowment points. The moral reference point r^b for the budget set represented by the flatter budget line is more favorable to the dictator than the moral reference point r^a for the set represented by the steeper budget line.

[FIGURE 7 ABOUT HERE: MMA and WARP]

Now consider two choices A and B from the original sets that violate the weak axiom of revealed preference (WARP). Suppose that the dictator chooses A on the steeper budget line. Then MMA (see Proposition 1) requires that A also be chosen from the quadrilateral set because it is a contraction of the feasible set represented by the steeper line that preserves the moral reference point. Suppose that B is chosen from the lower flat triangle. MMA requires that the choice in the quadrilateral (which is also a contraction of the lower flat triangle) allocates to the dictator less than B does, because r^a is to the left of r^b. But this contradicts the choice of A from the quadrilateral set. Thus, any pair of choices of type A and B that violate WARP also violate MMA. In fact, MMA places tighter restrictions on data than does WARP (e.g., in Figure 7 WARP implies point A must be northwest of the intersection whereas MMA implies it must be west of point B).

6.2 Give and Take: MMA vs. Warm Glow

Korenok et al. (2014) report a dictator game experiment to test the theoretical model of warm glow developed by Korenok et al. (2013). In particular, the authors explore the effects of changing

endowments and framing actions as giving to or taking from the recipient. Korenok et al. (2014) explain that data from their experiment is inconsistent with the predictions of their theory which, in this instance, are the same as the predictions of the conventional rational choice model. We have explained above that their data are consistent with MMA.[25]

6.3 MMA and Bully Games

MMA predicts both dictator game choices and social norms elicited by Krupka and Weber (2013). In their experiment, the moral reference point is (5, 0) in the standard dictator game and (2.5, 0) in the bully dictator game. Hence, MMA requires choices in the bully treatment to be drawn from a distribution that is less favorable to the dictator than the distribution of choices in the standard game. Therefore, we expect a higher amount allocated to the recipient and a positive estimate of the bully treatment in an ordered logistic regression. The reported mean amounts allocated to the recipients are $2.46 (standard) and $3.11 (bully) and the coefficient estimate for the bully treatment is significantly positive (see their Table 2).

Moreover, the distribution of elicited norms reported in Krupka and Weber's Table 1 are also consistent with MMA. A paired t-test of the two distributions rejects the null hypothesis of no effect (implied by CCA) in favor of the MMA-consistent alternative (approval of higher allocations to recipients). Hence, both actual choices and elicited beliefs in Krupka and Weber (2013) are consistent with MMA and highlight the importance of moral reference points.

6.4 MMA and Outside Options

Lazear, et al. (2012) report an extended experimental design for dictator games that includes an outside option that allows subjects to opt out of the dictator game. Their Experiment 1 is a between-subjects design in which one group of subjects plays a "distribute $10" dictator game and another group of subjects can choose an outside option, that pays the dictator $10 and the other subject $0, or choose to play the distribute $10 dictator game.[26] The Lazear, et al. Experiment 2 is a within-subjects design including several decisions with one selected randomly for payoff. In Decision 1,

---

[25] MMA is also consistent with the results from the meta-analysis in Zhang and Ortmann (2014) who find that the introduction of a take option leads to lower final payouts for the recipient. We should note however, that Dreber et al., (2013) report aggregate data patterns across give and take versions of the dictator game that appear to be at odds with the predictions of MMA. However, as noted in Zhang and Ortmann (2014, fn. 9) the analysis in Dreber et al. (2013) relies upon a normalized metric of sharing/giving that codes transfers in the Take only treatment as positive instead of negative. Further, it is important to note that there is imbalance in key demographics such as gender and age across treatments in Dreber et al. (2013). Since such factors have been shown to influence the amount a dictator is willing to share, it is not clear how to interpret differences in the normalized amount shared with the recipient in their data.

[26] In sessions run in Barcelona the pie was €10 while sessions in Berkeley used a $10 pie. The text of the paper uses the subject decision task description as an assignment to "divide $10 (€10)" while the subject instructions use the wording "distribute $10 (€10)".

subjects play a distribute $10 dictator game. In Decision 2, subjects can sort out of the $10 dictator game, and be paid $10 (with the other subject getting $0), or sort in and play the distribute $10 dictator game. In other decision tasks, subjects can sort out of a S dictator game, and be paid $10 (with the other subject getting $0), or sort in and play the distribute S dictator game. Values of S varied from 10.50 to 20.[27]

Explaining behavior of subjects in Experiment 2 who sorted into a $S > 10$ dictator game and kept more than 10 for themselves is straightforward. A more interesting behavior is that many subjects sorted out, and were paid 10, when they could have sorted into a $S > 10$ dictator game and retained more than 10 for themselves (and/or more than 0 for the other). For example, in the $S = 11$ game, the outside option pays (dictator, other) payoffs (10,0) whereas Pareto-dominating payoffs such as (11,0), (10.50, 0.50) and (10,1) are available to a subject who sorts into the dictator game. The reluctant/willing sharers model developed by Lazear et al. (2012) is consistent with behavior patterns in the experiment. That model is a utility function with three arguments: own payoff, other's payoff, and a binary indicator variable with value 1 for the sharing (dictator game) environment and value 0 for the non-sharing (outside option) environment. This type of behavior is consistent with our moral cost model in which choosing the outside option allows the decision maker to avoid moral costs from making the sharing decision whereas choosing to play the game involves this cost, as we now explain.

A subject has the right to choose the ordered pair of payoffs (10,0) by sorting out. This provides a clear endowment for the two-step game that includes the option of sorting in and paying the moral cost of making a sharing decision. Let S_j denote that amount of money that can be distributed in treatment j. Since the dictator's sharing options include 0 and S_j, the minimal expectations point for the two-stage game is the natural origin. Hence the moral reference point if the player sorts in is $(r_1, r_2) = (\frac{1}{2} \times 10, 0)$. Let preferences consistent with MMA be represented by a utility function $u(m - r_1, y - r_2)$. Substituting the budget constraint $m = S_j - y$ and the moral reference point (5,0) the decision problem for our agent becomes $\max_y u(S_j - y - 5, y)$. The MMA model is consistent with behavior by an agent who chooses the (10,0) outside option rather than sorting in to play a distribute $S > 10$ dictator game with feasible payoffs that Pareto-dominate (10,0) contained in its budget set.

---

[27] The experiment included anonymity and no-anonymity treatments.

Here we provide an example using a simple utility function, $u(m, y) = m + \gamma \sqrt{y}$. By sorting out, a subject can avoid the moral cost of making the sharing decision, obtain payoff allocation (10,0), and utility $V(out) = 10 + \gamma \times 0$. If the player sorts in then she incurs moral cost of making the sharing decision, instantiated in the model by the moral reference point $(r_1, r_2) = (5,0)$ and MMA. The decision-maker's optimization problem for the dictator game is

$$\max_{y \in [0,S]} u(m - r_1, y - r_2) = \max_{y \in [0,S]} (S - y - 5 + \gamma \sqrt{y}).$$

The optimal choice is $y^o = \gamma^2 / 4$ and the value of sorting in is $V(in) = S - 5 + \gamma^2 / 4$. Comparing it to the value of sorting out, $V(out) = 10$, one has:

1. Any agent with (*) $\gamma^2 < 4(15 - S)$ prefers sorting out and realizing payoff (10,0) to sorting in and being able to choose Pareto-dominating payoffs.

2. As S increases, inequality $S - 5 + \gamma^2 / 4 > 10$ becomes more likely to be satisfied and therefore the fraction of subjects sorting in increases, as observed in Experiment 2.

Experiment 1 in Lazear et al. (2012) is a between-subjects design in which one group of subjects play a distribute $10 dictator game and another group of subjects can sort out of the $10 dictator game, and be paid $10, or sort in and play the distribute $10 dictator game. The extended game with the outside option is modeled as above with the MMA model using the unambiguous (10,0) endowment provided by the outside option. The distribute $10 dictator game without outside options is a commonly used protocol for dictator games in which neither the dictator nor the recipient has a clearly assigned property right. This form of dictator game protocol is widely viewed as appropriate for research on sharing behavior but it does have an ambiguous endowment, as explained by Hoffman et al. (1994) and Hoffman, McCabe, and Smith (1996).[28] Experiment 1 data are consistent with predictions from the MMA model which follow from interpreting the 10 available for distribution as endowments to the dictator and recipient of $(10 - z, z)$, with $z > 0$.

6.5 MMA and Earned Endowments

Oxoby and Spraggon (2008) report an experiment that includes treatments whereby initial endowments are determined in a first stage. In the receiver earnings treatment, the recipient determined the initial endowment by their performance on a test that used 20 questions pulled from the Graduate Management Admissions Test (GMAT) or the Graduate Record Examinations

---

[28] The exact wording in the Hoffmann et al. subject instructions is "divide $10". The exact wording in the Lazear, et al. subject instructions is "distribute the $10 (€10)" although the text uses the wording "divide $10 (€10)".

(GRE). Depending upon the number of questions answered correctly, the recipient was provided an initial endowment of either CAN $10, CAN $20, or CAN $40. In the second stage, the dictator decided how much of this endowment they would like to take from the recipient. The dictator earnings treatment differed along two dimensions. First, the initial endowment was earned by the dictator's performance on the 20 question exam. Second, the dictator's decision in the second stage was to determine how much of the initial endowment they would like to give to the recipient.

Across both versions of the game, the minimal expectations point is (0, 0). Therefore, as in the Korenok et al. (2014) experiment, changes in moral reference points across the two treatments are entirely determined by changes in endowment. Focusing on pairs for whom the initial endowment is CAN $40, the moral reference point is (0,0) in the receiver-earning treatment and (20, 0) in the dictator-earning treatment. MMA would thus predict that the amount allocated to the recipient under the recipient earnings treatment is greater than the amount allocated to the recipient under the dictator earnings treatment. Across all three wealth levels, the mean amounts allocated to recipients in the receiver-earning treatment are greater than the mean amounts allocated to recipients in the dictator-earning treatment, which is a pattern of results at odds with CCA but consistent with the predictions of MMA.

Korenok et al. (2017) extend this line of inquiry by adding a set of survey questions designed to elicit participants' feelings of ownership over the initial endowments. As in Oxoby and Spraggon (2008), treatments varied whether the initial endowment was earned by the recipient or dictator and the subsequent framing of the task as either give to or take from the recipient. Across all wealth levels, the mean amount allocated to the recipient under the recipient earnings treatment was greater than the amount allocated to the recipient under the dictator earnings treatment. Moreover, dictators felt a stronger sense of ownership over the endowment than did recipients in the dictator earnings treatment and vice versa in the receiver earnings treatment. Hence, both actual choices and feelings of ownership over endowments depend on property rights and initial allocations. Such patterns are consistent with MMA and highlight the importance of moral reference points.

7. IMPLICATIONS OF MMA FOR PLAY IN GAMES WITH CONTRACTIONS

We next extend our discussion to illustrate the implications of MMA for play of strategic games involving contractions. Games that have been studied in previous literature include: (1) the moonlighting game and its contraction, the investment game, (2) carrot and stick games and a contraction in the positive domain (carrot game) as well as a contraction in the negative domain,

(stick game). Together with dictator games, these games have been widely used in the literature to measure different aspects of social behaviors, including trust and cooperation. MMA has different implications for play of these games than does CCA or a stronger traditional assumption such as convex preferences.

7.1 Investment and Moonlighting Games

The investment game (Berg, et al. 1995, and hundreds of other papers) can be constructed from the moonlighting game (Abbink, et al. 2000, and scores of other papers) by contracting the feasible choice sets of the first and second movers.[29] CCA and MMA have different implications for such contractions and allow a way to distinguish between the two models using observed choice.

First, we argue that, for any given *positive* amount received, the second mover's (SM's) choice is the same in the moonlighting and investment Games (with the same initial endowments). This is the prediction of CCA as well as MMA because the reference point for the SM opportunity sets is the same in the two games.

Next, we argue that for any first mover (FM) who sends a non-negative amount in the moonlighting game, CCA requires that he choose the same amount to send in the investment game. MMA, in contrast, requires him to choose a larger amount to send in the investment game. The reason for this difference is that the moral reference point for the FM opportunity set is more favorable to the FM in the moonlighting game than in the investment game.

An implication of the two statements is that MMA predicts more money being sent by all FMs in the investment game than in the moonlighting game whereas CCA makes this prediction only for FMs who take in the moonlighting game. Yet it is important to note that this latter "prediction" results solely from the constraint that prevents taking in the investment game, not from agent preferences in and of themselves.

Let e denote the endowment of each FM and each SM. The amount sent by the FM is denoted by s. If s is positive it is multiplied by $k > 1$ to obtain the amount received by the SM. Taking is not feasible in the investment game. In the moonlighting game, if s is negative then the multiplier is 1 to obtain the amount taken from the SM. The amount returned by the SM is denoted

---

[29] In the moonlighting game (Abbink, et al. 2000), both players are endowed with the same amount of money. The first mover (FM) can give or take money from the second mover (SM); the maximum amount that can be given is the full endowment whereas the maximum amount that can be taken is one-half the endowment. Money given by FM is tripled by the experimenter but money taken is not transformed. After the SM is informed of the FM's choice, he/she can also give or take money from the FM. Each currency unit (CU) taken costs SM 1/3 CU whereas each CU given costs SM one CU. The investment game is a contraction in that FM and SM can only give and not take.

by r. Returning a negative amount is not feasible in the investment game. In the moonlighting game, when r is negative it costs the SM r/k to take r from the FM.

SM opportunity sets across the two games: Let the SM be in information set M_s for some *non-negative* amount s sent by the FM in the moonlighting game. The M_s set contains costly options for the SM but can increase/decrease FM's monetary payoff: $M_s = M_s^+ \cup M_s^-$ where

$$M_s^+ = \left\{(e-s+r, e+ks-r) : r \in [0, ks]\right\}$$
$$M_s^- = \left\{(e-s+r, e+ks+r/k) : r \in [-(e-s)/k, 0)\right\}$$

Consider the SM's choice in M_s in the moonlighting game when the FM sends a non-negative amount. Consistent with observed behavior[30] (as well as Pareto efficiency), the amount returned will be from M_s^+.

What are CCA and MMA predictions for SM's choice in the investment game, at information set I_s given the same nonnegative s? In the investment game the SM's choices can only increase the FM's monetary payoff by decreasing own monetary payoff,

$$I_s = \left\{(e-s+r, e+ks-r) : r \in [0, ks]\right\}$$

Thus $I_s = M_s^+ \subset M_s$. CCA requires the same $r_s \in M_s^+$ to be the SM's choice in the investment game. This is also the MMA prediction because sets M_s and I_s have the same moral reference point, with coordinate $e-s$ for the FM and $e+ks/2$ for the SM.

FM choices across the two games: In the moonlighting game, the FM can send money to the SM or take up to one-half of the SM's initial endowment. Any positive amount sent ($s > 0$) is multiplied by $k > 1$. Any amount taken ($s < 0$) is not transformed (it is one for one). The FM choice set is $M = M^+ \cup M^-$ where

$$M^+ = \{(e-s, e+ks) : s \in [0, e]\}$$
$$M^- = \{(e-s, e+s) : s \in [-e/2, 0)\}$$

---

[30] Only 2 (out of 46) second movers who did not have money taken from them by first movers choose $r_s \in M_s^-$.

Suppose the FM's choice in the moonlighting game is some non-negative s_M. In the investment game, the FM can only send money to the SM. So, $I = M^+ \subset M$ as the FM choice set is

$$I = \{(e-s, e+ks) : s \in [0,e]\}$$

CCA requires the non-negative amount s_M to be the FM's choice in the investment game when it is the choice in the moonlighting game because the feasible set in the investment game is a contraction of the feasible set in the moonlighting game. In contrast, MMA implies that the FM will send more in the investment game because the moral reference point, $(e/2, e)$ in set I is more favorable to the SM than is the moral reference point $(e/2, e/2)$ in set M.

Implications for game play: Both CCA and MMA imply that, for any *positive* amount received, the SM's choices in the moonlighting and investment games are identical. We distinguish between two types of FMs: the ones who send in the moonlighting game and the ones who take. For a FM who takes in the moonlighting game, by design of the two games the FM must send more in the investment game. For a FM who does not take in the moonlighting game, we have shown above that CCA predicts the same amount being sent in the two games whereas MMA predicts a larger amount being sent in the investment game.

Existing data that provide empirical support for MMA: We have analyzed data from an investment game experiment reported in Cox (2004) and a moonlighting game experiment reported in Cox, Sadiraj, and Sadiraj (2008). These two experiments used the same initial endowments $e=(10,10)$, the same multiplier k (=3) and were run by the same experimenter. Data from these experiments are consistent with the implications of MMA and inconsistent with the implications of CCA, as follows.

We have data from 64 subjects who participated in the investment game and 130 subjects (66 within-subjects design and 64 between-subjects design) who participated in the moonlighting game.

FM choices: Using only FM data with non-negative amounts sent, we find that the means of the amounts sent are 5.97 (IG) and 4 (MG) and significantly different (t-test, p-value= 0.026.[31] Therefore the FM data are consistent with the above implications of MMA but inconsistent with implications of CCA.

---

[31] If we look only at Send > 0, means are 7.35 (IG) and 4.84 (MG) and significantly different (t-test, p-value=0.004).

SM choices: Estimates (standard errors in parentheses) of censored regressions for SM choices at information sets with "FM not taking" (send ≥ 0, N=78) are[32]

$$E(r^s) = 0.67^{***}(\pm 0.15) \times s + 0.41(\pm 0.29) \times s \times D_M - 0.23(\pm 1.30) \times D_M$$

Insignificance of the coefficients for D_M and $s \times D_M$, "Moon" and "Send*Moon," are consistent with the (same) implication of MMA and CCA, as discussed above.

Taken jointly, we conclude that differences in play across the moonlighting and investment games are inconsistent with standard rational choice theory. Changes in the first mover's moral reference points across games leads to greater amounts shared in the investment game, a finding that is consistent with the predictions of MMA.

7.2 Carrot, Stick, and Carrot/Stick Games

Andreoni, Harbaugh and Vesterlund (2003) look at effects of rewards and punishments on cooperation by studying behavior in three games: the carrot game that offers incentives only in terms of rewards, the stick game that allows only for negative incentives (punishment) and the carrot and stick (C&S) game that offers players both types of incentives. The two single incentive games are natural contractions of the C&S game. We argue that for any given positive amount received the SM's predicted choice is *the same* in the C&S and carrot game. This is the prediction of CCA as well as MMA and arises as the moral reference point of the SM's opportunity set is the same in the two games. Next, we argue that for any positive amount received the SM's predicted choice is *less malicious* in the stick game than in the C&S game according to MMA because the moral reference point in the stick game favors the SM.

Let $e = (240,0)$ in cents denote the endowments of the FM and the SM. The amount sent, s by the FM is the amount received by the SM and can take values from [40, 240] in all three games. The return, r_s by the SM can be positive (carrot), negative (stick) or either (C&S game) as returning a negative amount is not feasible in the carrot game whereas returning a positive amount is not feasible in the stick game. Despite the sign of the amount returned, the FM receives $5r_s$.

SM choices across the three games: For the amount s sent by the FM let the SM feasible sets be denoted by M_{cs}^s in the C&S game, M_c^s in the carrot game and M_s^s in the stick game such

---

[32] Send > 0 (N=64): $E(r^s) = 0.65^{***}(\pm 0.17) \times s + 0.42(\pm 0.36) \times s \times D_M - 0.14(\pm 1.87) \times D_M$

that $M_{cs}^s = M_c^s \bigcup M_s^s$. The M_{cs}^s set consists of options that are all costly for the SM but can increase or decrease FM's monetary payoff. The sets are:

$$M_c^s = \{(240 - s + 5r, s - r) : r \in [0, s]\}$$
$$M_s^s = \{(240 - s + 5r, s + r) : r \in [\max\{-(240 - s)/5, -s\}, 0]\}$$

Let r_{cs} be the SM's choice in the C&S game when the FM sends amount s. CCA and MMA predictions for SM's choice when the FM sends amount s are as follows:

a. Carrot game: In this game the SM's choices can only increase the FM's monetary payoff by decreasing own monetary payoff. CCA requires that if the SM choice in the C&S game is positive, i.e. $r_{cs} \in M_c^s$ then it remains a most preferred return in the carrot game. This is also the MMA prediction because sets M_{cs}^s and M_c^s have the same moral reference point, $(240 - s)$ as the FM coordinate and $(s/2)$ as the SM coordinate. Andreoni et al. (2003, Figure 7) find larger demand for rewards in the C&S game than in the carrot game which is inconsistent with both CCA and MMA.

b. Stick game: In this game the SM's choices can only decrease the FM's monetary payoff by decreasing own monetary payoff. CCA requires that if the SM's most preferred choice in the C&S game is to reduce the FM's monetary payoff, i.e., $r_{cs} \in M_s^s$ then it remains a most preferred return in the stick game. MMA, however, predicts in the stick game a smaller return in absolute value because the moral reference point favors the SM as its coordinate is s (rather than $s/2$) whereas the FM's coordinate remains the same, $(240 - s)$. Andreoni et al. (2003, Figure 6) report a result they characterize as "surprising" (pg. 898) that demand for punishment is larger in the C&S game than in the stick game. This result is predicted by MMA but is inconsistent with CCA.

Taken in its totality, data from Andreoni et al. (2013) provides evidence inconsistent with standard rational choice theory and mixed support for MMA. Importantly, however, MMA can rationalize a data pattern that Andreoni et al. (2013) label as surprising, that the demand for punishment is greater in the C&S game than in the stick game. As the moral reference point for the SM in the stick game is more favorable than in the C&S game, which is what one would expect under MMA.

8. CONCLUDING REMARKS

When faced with the opportunity to share resources with a stranger, when and why do we give? The dictator game has emerged as a key data generator to provide researchers with a simple

approach for eliciting other-regarding preferences in a controlled setting. The game has worked well in the sense that we now understand giving behaviors at a much deeper level. What has been less well explored is whether received results violate the basic foundations of economic theory.

As we explain, recent dictator game experiments reveal that choices of subjects in specific pairs of dictator games are inconsistent with convex preference theory (List, 2007; Bardsley, 2008; Cappelen et al., 2013) and inconsistent with (more general) rational choice theory (Korenok, et al., 2014) characterized by the Contraction Consistency Axiom (CCA).

The designs of experiments that produce the anomalous data suggest how to extend rational choices theory to increase its empirical validity. The Korenok, et al. (2014) experimental design and data suggest that choices depend on endowment in ways not captured by conventional theory. The List (2007) and Bardsley (2008) experimental designs and data suggest that choices depend on minimum and maximum feasible payoffs in ways not captured by conventional theory. In this spirit, we propose moral reference points and a Moral Monotonicity Axiom (MMA) that models dependence on endowment and minimal expectations payoffs. An implication of MMA is preservation of the contraction property of rational choice theory for feasible sets and subsets that have the same moral reference point. The moral reference points we propose are observable features of feasible sets, not subjective reference points that can be adjusted *ex post* to fit new data.

We report an experiment designed to test the central feature of the new model: monotonicity in choice with respect to distinct dimensions of observable moral reference points. Data from the experiment imply rejection of CCA in favor of MMA.

The MMA model, however, has more general applicability. We explain how it can rationalize data from other types of dictator games in the literature. We also explain how the model has implications for play of strategic games involving contractions of feasible sets that differ from implications of conventional theory.

The model and experimental data lead us to conclude that moral reference points play a major role in the decision to act generously. As a whole, these findings highlight the importance of revisiting standard models to explore the role of moral reference points in a broader array of choice settings. In the paper, we have provided an explanation of how the theory of moral reference points is predictive of received findings in a range of economic games designed to elicit social and cooperation behaviors. In this manner, we view our results as having both positive and normative import. For empiricists and practitioners, the results herein provide an indication that moral costs can play an important role in welfare calculations and program evaluation.

REFERENCES

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner, "The Moonlighting Game: An Empirical Study on Reciprocity and Retribution." *Journal of Economic Behavior and Organization* 42, (2000): 265-77.

Afriat, Sidney N., "The Construction of Utility Functions from Expenditure Data," International Economic Review, 8, no.1 (1967), 67–77.

Andersen, Steffen, Seda Ertaç, Uri Gneezy, Moshe Hoffman, and John A. List. "Stakes Matter in Ultimatum Games." *American Economic Review* 101, no. 7 (2011): 3427-3439.

Andreoni, James, and John H. Miller. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70, no. 2 (2002): 737-753.

Andreoni, James, William Harbaugh, and Lise Vesterlund. "The Carrot or the Stick: Rewards, Punishments, and Cooperation." *American Economic Review* 93, no. 3 (2003): 893-902.

Arrow, Kenneth J. "Rational Choice Functions and Orderings." *Economica* 26, no. 102 (1959): 121-127.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. "Social preferences and the response to incentives: Evidence from personnel data." *The Quarterly Journal of Economics* 120, no. 3 (2005): 917-962.

Bardsley, Nicholas. "Dictator Game Giving: Altruism or Artefact?" *Experimental Economics* 11, no. 2 (2008): 122-133.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10, no. 1 (1995): 122-42.

Bolton, Gary E., and Axel Ockenfels. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90, no. 1 (2000): 166-193.

Cabrales, Antonio, Raffaele Miniaci, Marco Piovesan, and Giovanni Ponti. "Social preferences and strategic uncertainty: an experiment on markets and contracts." *The American Economic Review* 100, no. 5 (2010): 2261-2278.

Cappelen, Alexander W., Ulrik H. Nielsen, Erik Ø. Sørensen, Bertil Tungodden, and Jean-Robert Tyran. "Give and Take in Dictator Games." *Economics Letters* 118, no. 2 (2013): 280-283.

Çelen, Boğaçhan, Andrew Schotter, and Mariana Blanco. "On blame and reciprocity: Theory and experiments." *Journal of Economic Theory* 169 (2017): 62-92.

Charness, Gary, and Matthew Rabin. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117, no. 3 (2002): 817-869.

Chernoff, Herman. "Rational Selection of Decision Functions." *Econometrica* 22, no. 4 (1954): 423-443.

Choi, Syngjoo, Douglas Gale, Ray Fisman, and Shachar Kariv. "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review* 97, no. 5 (2007): 1921-1938.

Cox, James C. "How to Identify Trust and Reciprocity." *Games and Economic Behavior* 46, no. 2 (2004): 260-281.

Cox, James C., Daniel Friedman, and Vjollca Sadiraj. "Revealed Altruism." *Econometrica* 76, no. 1 (2008): 31-69.

Cox, James C., and David M. Grether. "The Preference Reversal Phenomenon: Response Mode, Markets and Incentives." *Economic Theory* 7, no. 3 (1996): 381-405.

Cox, James C., John A. List, Michael Price, Vjollca Sadiraj, and Anya Samek. *Moral costs and rational choice: Theory and experimental evidence*. Working Paper No. w22234. National Bureau of Economic Research, 2016.

Cox, James C., Klarita Sadiraj and Vjollca Sadiraj. "Implications of Trust, Fear, and Reciprocity for Modeling Economic Behavior." *Experimental Economics* 11 (2008): 1-24.

Cox, James. C., and Vjollca Sadiraj. "On Modeling Voluntary Contributions to Public Goods." *Public Finance Review* 35, no. 2 (2007): 311-332.

Cox, James C., and Vjollca Sadiraj. "A Theory of Dictators' Revealed Preferences." Paper presented at the Biennial Social Dilemmas Conference, Rice University (2010).

Cragg, John G. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica* 39, no. 5 (1971): 829-844.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127, no. 1 (2012): 1-56.

Dreber, A., T. Ellingsen, M. Johannesson, and D. G. Rand. "Do people care about social context? Framing effects in dictator games." *Experimental Economics*, 16(3), (2013): 349-371.

Dufwenberg, Martin, and Georg Kirchsteiger. "A Theory of Sequential Reciprocity." *Games and Economic Behavior* 47, no. 2 (2004): 268-298.

Engel, Christoph. "Dictator Games: A Meta Study." *Experimental Economics* 14, no. 4 (2011): 583-610.

Engelmann, Dirk, and Guillaume Hollard. "Reconsidering the Effect of Market Experience on the "Endowment Effect"." *Econometrica* 78, no. 6 (2010): 2005-2019.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics* 108, no. 2 (1993): 437-459.

Fehr, Ernst, and Klaus M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114, no. 3 (1999): 817-868.

Ferraro, Paul J., and Michael K. Price. "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment." *Review of Economics and Statistics* 95, no. 1 (2013): 64-73.

Fisman, Ray, Shachar Kariv, and Daniel Markovits. "Individual Preferences for Giving." *American Economic Review* 97, no. 5 (2007): 1858-1876.

Fisman, Ray, Pam Jakiela, and Shachar Kariv. "How Did the Great Recession Impact Social Preferences." *Journal of Public Economics* 128 (2015): 84-95.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* 40, no. 2 (2002): 351-401.

Fudenberg, Drew, and David K. Levine. "Fairness, Risk Preferences and Independence: Impossibility Theorems." *Journal of Economic Behavior & Organization* 81, no. 2 (2012): 606-612.

Galperti, Simone, and Bruno Strulovici. "A Theory of Intergenerational Altruism." *Econometrica* 85, no 4 (2017): 1175-1218.

Grossman, Philip J. and Catherine C. Eckel, "Giving vs. Taking for a Cause." *Economics Letters* 132 (2015): 28-30.

Harless, David W. "Predictions about Indifference Curves inside the Unit Triangle: A Test of Variants of Expected Utility Theory." *Journal of Economic Behavior & Organization* 18, no. 3 (1992): 391-414.

Herz, Holger and Dmintry Taubinsky. "What Makes a Price Fair? An Experimental Study of Transaction Experience and Endogenous Fairness Views." *Journal of the European Economic Association* (2017), forthcoming.

Hey, John D., and Chris Orme. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica* 62, no. 6 (1994): 1291-1326.

Hicks, John R. *Value and Capital*. Oxford: Clarendon Press. (1946).

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. "Preferences, property rights, and anonymity in bargaining games." *Games and Economic Behavior* 7, no. 3 (1994): 346–380.

Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith. 1994. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review* 86, no. 3 (1996): 653-660.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. "Fairness and the Assumptions of Economics." *Journal of Business* 59, no. 4 (1986): 285-300.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* 98, no. 6 (1990): 1325-1348.

Kessler, Judd B. and Leider, Stephen. "Norms and Contracting." *Management Science* 58, no. 1 (2012): 62–77.

Kimbrough, Erik O., and Alexander Vostroknutov. "Norms Make Preferences Social." *Journal of the European Economic Association*, (2015): Forthcoming.

Korenok, Oleg, Edward L. Millner, and Laura Razzolini. "Impure Altruism in Dictators' Giving." *Journal of Public Economics* 97, (2013): 1-8.

Korenok, Oleg, Edward L. Millner, and Laura Razzolini. "Taking, Giving, and Impure Altruism in Dictator Games." *Experimental Economics* 17, no. 3 (2014): 488-500.

Korenok, Oleg, Edward L. Millner, and Laura Razzolini. "Feelings of Ownership in Dictator Games." *Journal of Economic Psychology* 61, (2017): 145-151.

Kőszegi, Botond, and Matthew Rabin. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics* 121, no. 4 (2006): 1133-1165.

Krupka, Erin L. and Roberto A. Weber. "Identifying Social Norms using Coordination Games: Why does Dictator Game Sharing vary?" *Journal of the European Economic Association* 11, no. 3 (2013): 495-524.

Laibson, David. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112, no. 2 (1997): 443-478.

Landry, Craig E., Andreas Lange, John A. List, Michael K. Price, and Nicholas G. Rupp. "Is a Donor in Hand Better Than Two in the Bush? Evidence from a Natural Field Experiment." *American Economic Review* 100, no. 3 (2010): 958-983.

Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber. "Sorting in Experiments with Application to Social Preferences." *American Economic Journal: Applied Economics* 4, no. 1 (2012): 136-163.

Levitt, Steven D., and John A. List. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21, no. 2 (2007): 153-174.

List, John A. "Preference Reversals of a Different Kind: The "More Is Less" Phenomenon." *American Economic Review* 92, no. 5 (2002): 1636-1643.

List, John A. "Neoclassical theory versus prospect theory: Evidence from the marketplace." *Econometrica* 72, no. 2 (2004): 615-625.

List, John A. "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy* 115, no. 3 (2007): 482-493.

O'Donoghue, Ted, and Matthew Rabin. "Doing It Now or Later." *American Economic Review* 89, no. 1 (1999): 103-124.

O'Donoghue, Ted, and Matthew Rabin. "Choice and Procrastination." *Quarterly Journal of Economics* 116, no. 1 (2001): 121-160.

Oxoby, Robert J., and John Spraggon. "Mine and Yours: Property Rights in Dictator Games." *Journal of Economic Behavior and Organization* 65, no. 3-4 (2008): 703-713.

Rabin, Matthew. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83, no. 5 (1993): 1281-1302.

Samuelson, Paul A. "A Note on the Pure Theory of Consumer's Behavior." *Economica* 5, no. 17 (1938): 61-71.

Samuelson, Paul A. *Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press, 1947.

Sen, Amartya K. "Choice Functions and Revealed Preference." *Review of Economic Studies* 38, no. 3 (1971): 307-317.

Sen, Amartya K. "Social Choice Theory." in *Handbook of Mathematical Economics*, vol. III, ed. by K. J. Arrow and M. Intriligator. Amsterdam: North-Holland (1986).

Slovic, Paul. "The Construction of Preference." *American Psychologist* 50, no. 5 (1995): 364-371.

Tversky, Amos. "Intransitivity of Preferences." *Psychological Review* 76, no. 1 (1969): 31-48.

Varian, Hal R., "The Nonparametric Approach to Demand Analysis." *Econometrica* 50, no. 4 (1982): 945–973.

Zhang, Le, and Andreas Ortmann. "The Effects of Take-Option in Dictator-Game Experiments: A Comment on Engel's (2011) Meta-study." *Experimental Economics* 17, no. 3 (2014): 414-420.

Table 1. Summary Statistics

		Average Transfer	Average. Final Payoffs	Nobs
Inequality	Give	4.54 (2.96)	(22.46, 7.54)	61
	Take	4.19 (3.34)	(22.81, 7.19)	81
	Symmetric	6.12 (4.95)	(20.88, 9.12)	82
Equal	Give	2.65 (2.24)	(20.35, 9.65)	66
	Take	3.17 (2.88)	(19.83, 10.17)	58
	Symmetric	3.94 (3.52)	(19.06, 10.94)	62
Envy	Give	2.43 (1.65)	(16.57, 13.43)	67
	Take	2.06 (1.85)	(16.94, 13.06)	69
	Symmetric	2.64 (2.55)	(16.36, 13.64)	66

"Ave. Transfer" is the amount by which the average recipient's payoff chosen by dictators exceeds the recipient's minimum expectations payoff (standard deviations in parentheses).

Table 2. Tests for Effects of Recipient Moral Reference Dimension ($r_1 = 15$)

| Dep. Var: Recipient's Final Payoff	Hurdle Model		Tobit Model	
	(1)	(2)	(1)	(2)
$r_2\ [\geq 0]$	0.134	0.136	0.674	0.668
	(0.031)	(0.031)	(0.187)	(0.186)
constant			6.145	6.955
			(1.548)	(2.417)
Demographics	no	yes	no	yes
Log-likelihood	-247.28	-244.15	-261.28	-258.30
Observations	207		207	

Notes: MMA predicted sign in square brackets. Entries are average marginal effects (Hurdle Model) and coefficients (Tobit model). Lower and upper bounds in regressions are 11 and 15. Nr of observations: 107 (left-censored), 58 (right-censored) and 42 (un-censored). Standard errors in parentheses. Demographics include gender, race, GPA, religion, major and study year.

Table 3. Moral Reference Points and Transfers

Dep. Var: Transfer	Hurdle Model		Tobit Model	
	(1)	(2)	(1)	(2)
$r_1\ [\leq 0]$	-0.058	-0.055	-0.098	-0.104
	(0.027)	(0.027)	(0.047)	(0.047)
$r_2\ [\geq -1]$	-0.319	-0.314	-0.497	-0.487
	(0.047)	(0.047)	(0.091)	(0.090)
constant			7.918	7.733
			(1.301)	(1.547)
Demographics	no	yes	no	yes
Observations	612		612	

Notes: MA predicted sign in brackets. Entries are average marginal effects (Hurdle Model) and coefficients (Tobit model). Lower and upper bounds are 0 and 8. Nr of observations: 166 (left-censored), 95 (right-censored) and 351 (un-censored). Robust standard errors (clustered at subject ID level) in parentheses. Demographics include gender, race, GPA, religion, major and study year.

Figure 1. Histograms using Data from List (2007) and Bardsley (2008)
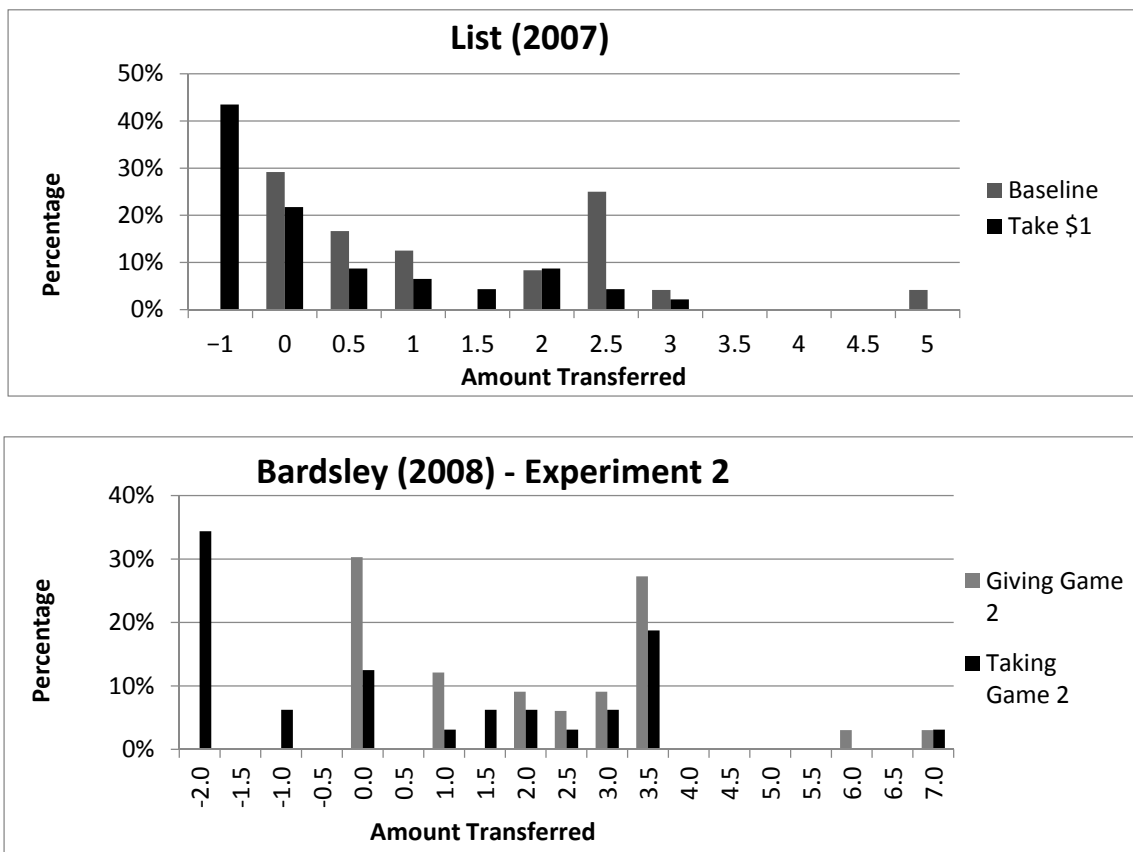
Notes: In the upper panel, Baseline refers to the standard dictator game in which dictators can choose to give $0 to $5 to the receivers. The Take $1 refers to the dictator game in which the feasible set is augmented to allow taking $1 from the recipient. In the lower panel, the Giving Game 2 refers to a standard dictator game in which dictators can choose to give $0 to $7 to receivers. Taking Game 2 refers to a game that is augmented to allow taking $2 from the recipient.

Figure 2. Example of Choices with non-Convex Preferences
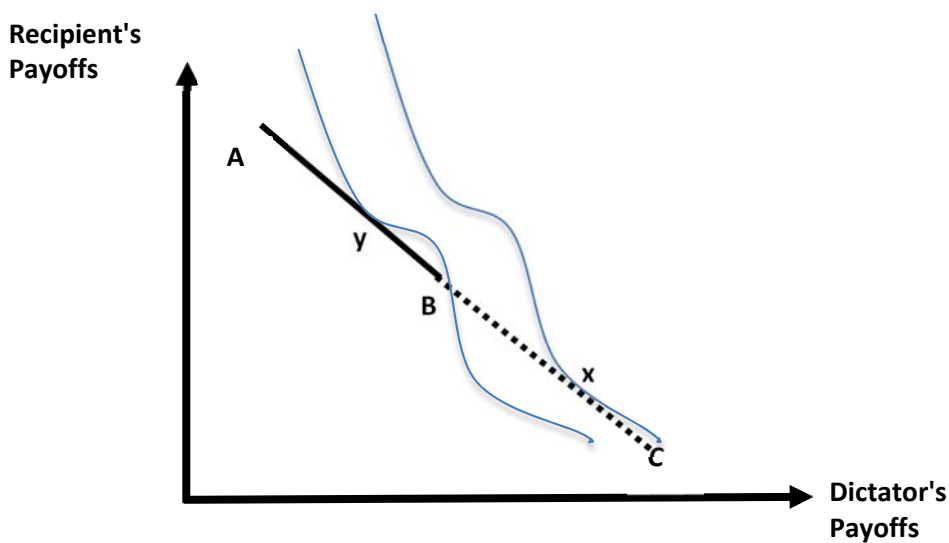
Figure 3. Endowments, Average Choices, and Moral Reference Points for Korenok, et al.
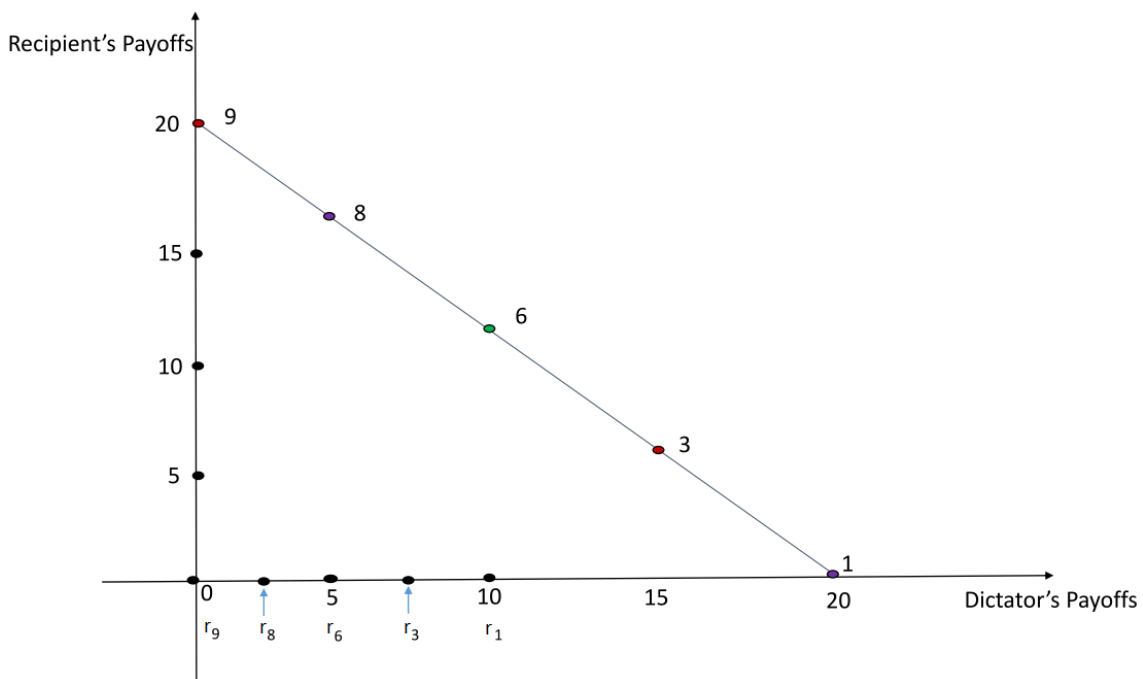

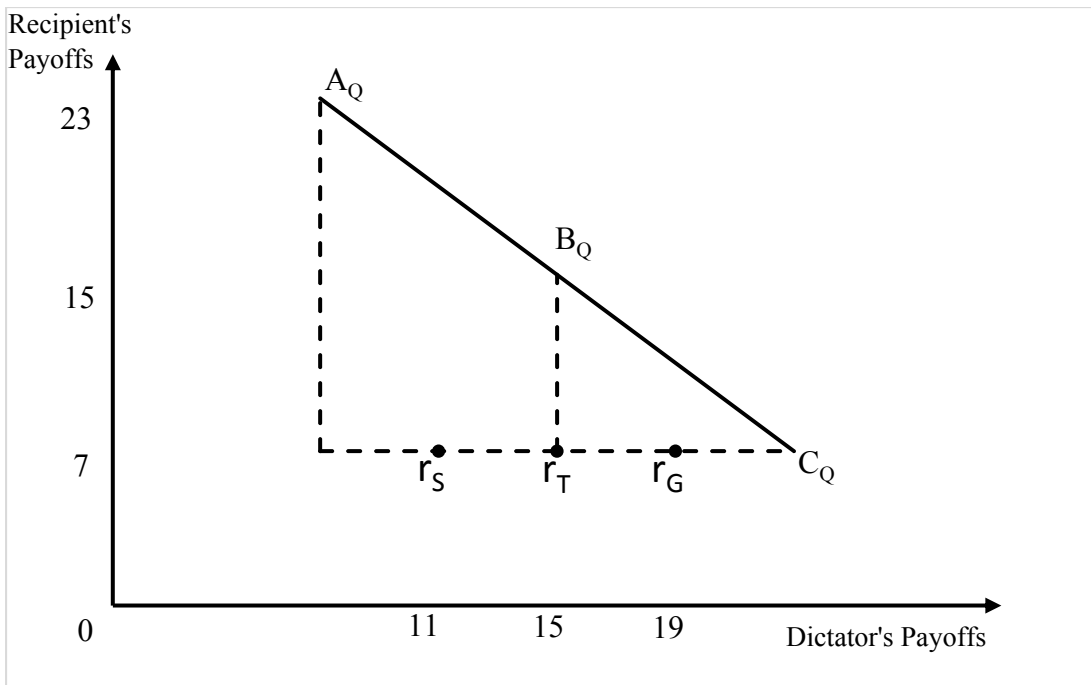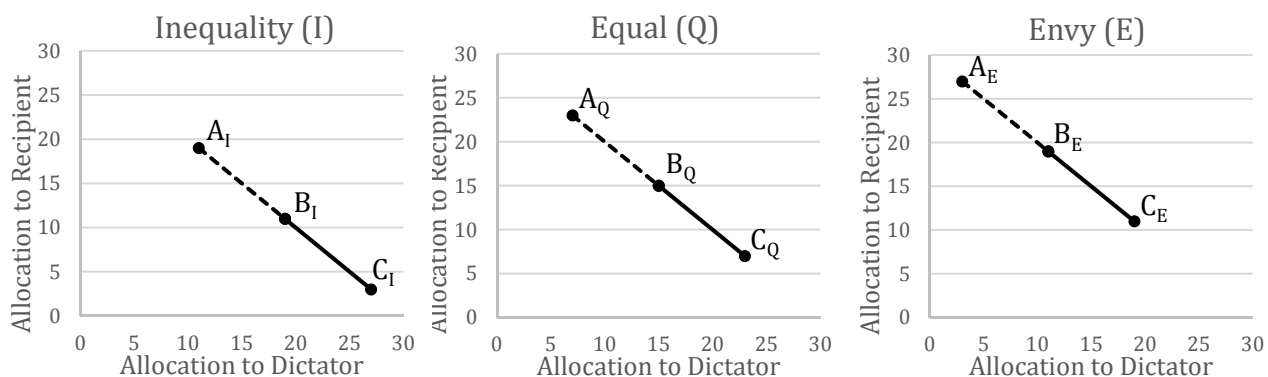
Figure 4. Moral Reference Points in Treatment Q

Figure 5. Feasible Sets: [B, C] for Give or Take, [A, C] for Symmetric

Notes: This figure portrays the feasible allocations for each treatment and action set. Participants in the Give or Take action sets can choose from [B, C], while participants in the Symmetric action set can choose from [A, C]. Actual feasible choices are ordered pairs of integers on the line segments.
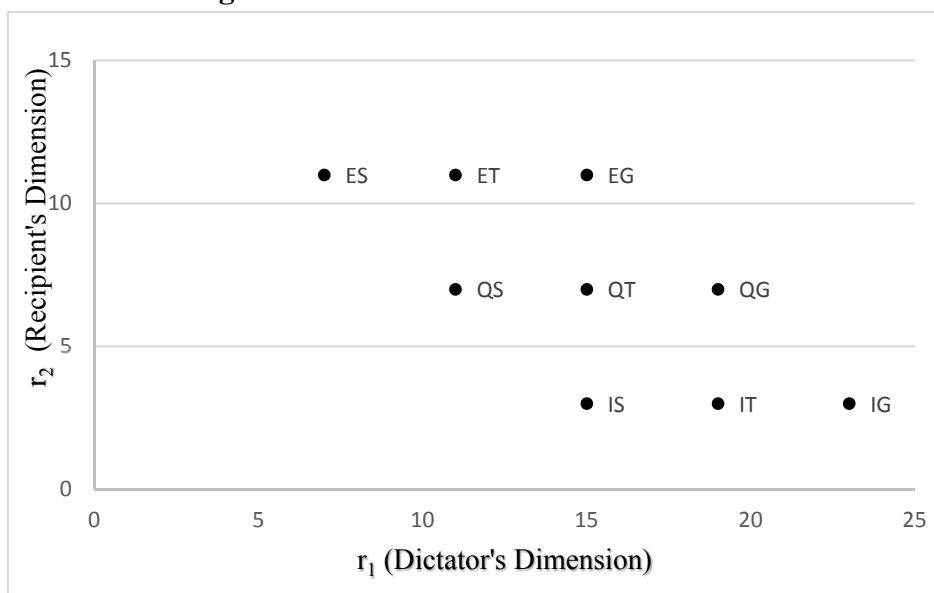
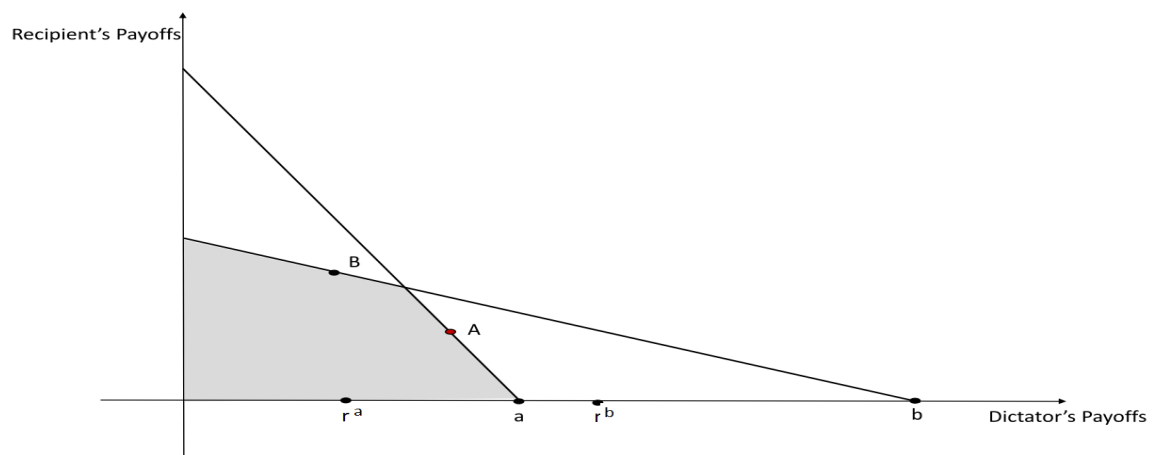**Figure 6. Moral Reference Points for Treatments**

Figure 7. MMA Implies WARP for the Andreoni and Miller Experiment

APPENDICES

Appendix A. Proof of Proposition 1

Let f belong to both F^* and G. Consider any g from G^*. As G and F have the same moral reference point, $r^g = r^f$, MMA requires that $g_i \geq f_i$ and $g_i \leq f_i$, $\forall i$. These inequalities can be simultaneously satisfied if and only if $g = f$, i.e. f belongs to G^* which concludes the proof for Property α_M. Note, though, that any choice g in G^* must coincide with f, an implication of which is G^* must be a singleton. So, if the intersection of F^* and G is not empty then choices satisfy Property β_M.

Appendix B. Moral Reference Point in the Presence of N Players

Endowments for n agents will typically be specified, hence are observable. Identification of observable minimal expectations payoffs for $n \geq 2$ players can proceed as follows. Let y denote the vector of payoffs of n players. Let the feasible set be a finite set F. Let y_j^o be the maximum feasible payoff for player j $(= 1, 2, \cdots n)$, that is

$$y_j^o(F) = \max\{y_j \mid y \in F\}$$

The minimal expectations point, y_*^F is defined as follows. For each player j, define player i's minimal expectation payoff with respect to j as

$$y_{*ij}^F = \min\{y_i \mid (y_{-j}, y_j^o) \in F\}$$

Let $S_i = \{y_{*ij}^F : j \neq i)$ be the set of i's minimal expectation points. Naturally, player i expects her payoff to be no smaller than the smallest element in S_i; thus $y_{*i}^F = \min S_i$, which is the ith element of the vector y_*^F.

Appendix C. Effect of Moral Reference Point on Transfers

Let "Transfer" be defined as the amount by which the recipient's payoff exceeds her minimum expectations payoff. In a Give treatment, the transfer is the amount the dictator gives to the recipient. In a Take treatment, the transfer is the amount the dictator does *not* take from the recipient. In a Symmetric treatment, the transfer is the amount not taken plus the amount given, if any. In all treatments, the dictator makes a choice of an amount to give or take that we here represent by a transfer, $t \in T$, where $T=[0,16]$ in a Symmetric treatment (Envy, Equal and

Inequality) and $T=[0,8]$ in the Give/Take scenarios (Envy, Equal and Inequality). The feasible set is

$$X = \left\{ (m,y) \mid m + y = 30, y = y_0 + t, t \in T \right\}$$

where (m,y) are dictator's and recipient's final monetary payoffs and y_0 is recipient's minimum feasible payoff. Let e and r be the initial endowment and the moral reference point of set X, that is, $r_1 = \frac{1}{2}(\max T + e_1)$ and $r_2 = y_0$. If the dictator chooses $t \in T$ then the recipient's and dictator's final payoff are $y = t + r_2$ and $m = 30 - y = 30 - (t + r_2)$.

CCA Choices: Let $P^* = (m^*, y^*)$ be the dictator's chosen allocation of \$30. Then when the dictator faces any subset of X that contains P^*, by CCA the dictator's choice of transfer t^o is such that $t^o + r_2 = y^*$. Thus, if r_2 increases then the chosen transfer, t^o decreases by the same amount for as long as the set X contains P^*. Next, preservation of y^* under CCA is not affected by r_1 as y^* does not depend on r_1. Thus we have the following hypothesis:

Hypothesis CCA: The chosen transfer, t^o is not affected by r_1 and $\Delta t^r / \Delta r_2 = -1$.

MMA Choices: Let dictator's choice satisfy MMA. One way to think about a dictator who is "socially" cautious is that he can claim social credits only for the transfer part, t^r rather than all of the recipient's payoff, $r_2 + t^r$ (because the recipient gets at least r_2 from the experimenter no matter what the dictator chooses).

By MMA, dictator's final chosen payoff increases in r_1. Because the total budget is fixed (at \$30), the dictator's payoff increasing in r_1 implies that the recipient's payoff is decreasing in r_1, which for fixed r_2 implies a decreasing transfer, t^r. It follows that the chosen transfer t^r must decrease in r_1. Next, by MMA, recipient's payoff, $r_2 + t^r$ increases in r_2, which implies $1 + \Delta t^r / \Delta r \geq 0$. Thus we have the following hypothesis:

Hypothesis MMA: The chosen transfer, t^r decreases in r_1 and $\Delta t^r / \Delta r_2 \geq -1$.