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, RWI-Leibniz Institute for Economic Research, and NBER
[d]University of Southern California

ABSTRACT

The literature exploring other-regarding behavior uncovers interesting phenomena, yet the extent the data subscribes to foundational assumptions of economics is not well understood. We explain how recent work challenges rational choice theory as well as its special case, convex preference theory. We propose a new axiom that modifies classic choice theory and exhibits choice monotonicity to observable reference points, establishing consistency with otherwise-anomalous data. We design experiments that provide a direct test of the new axiom. Data from our experiments and previous dictator experiments support the new axiom. We also apply the new axiom to several sequential strategic games.

JEL Classifications: C93, D03, D64
Keywords: rational choice, moral reference points, giving, taking, experiment

---

* We thank the John Templeton Foundation under the Science of Philanthropy Initiative and the National Science Foundation (grant number SES-1658743) for research funding. We thank Yongsheng Xu for helpful comments and suggestions.

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 fairness, equity, efficiency, 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; Herz 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 elicited using a class of experiments taking the form of dictator games, gift exchange games, public goods games, ultimatum games, and trust games. While such games have shown that social preferences touch many areas of economic interactions, the literature provides little guidance as to whether individual choices in such settings satisfy deeply held economic tenets.[1] The shortage of work exploring 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

---

[1] There is some evidence that certain 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. Fisman et al. (2007) extend Andreoni and Miller (2002) 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. More recently, Andersen et al. (2011) provide data that reveals demand curves for fairness in an ultimatum game are downward sloping.

function for each individual. These 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 work by exploring whether basic economic tenets are satisfied in sharing choices as observed in the economics literature. 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, does not challenge 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] Yet, convexity of preferences is not required by rationality, which for singleton choice sets, is equivalent to the Contraction Consistency Axiom (Sen, 1971).[4]

Building upon this discussion, we advance a new choice axiom of moral monotonicity 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

---

[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 (i.e., subset) of 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 interested reader should see data from one of the treatments in Korenok, et al. (2014), who provide tests of the Contraction Consistency Axiom.

accommodate dictator game data that violates a central tenet of conventional theory – in this case, 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 report on two experiments designed to test directly the new "moral" choice axiom – monotonicity in choice with respect to the dimensions that define moral reference points. Results from our experiment provide support for the new axiom that captures the observed patterns of sharing. In contrast, the data are at odds with the standard model of rational choice and with familiar models of preferences (referred to herein as "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 study is structured as follows. Section 2 explores the implications for choice in distinct types of dictator games in the previous literature that challenge: (a) *homo economicus* preferences; (b) other-regarding convex preferences; and (c) general consequentialist rational choice theory. Section 3 introduces the new axiom of moral monotonicity choice and Section 4 presents the design of our experiments that 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 Korenok et al. (2014), Krupka and Weber (2013), Lazear, Malmendier, and Weber (2012), Oxoby and Spraggon (2008) as well as for the classical (loss aversion) reference-dependent model of Tversky and Kahneman (1991). 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

[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). However, such models incorporate moral costs as parameterizations of a utility function. Our approach differs from this prior work by developing an axiom that extends choice theory to allow for moral reference points.

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 (Engel, 2011).

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, preferences may be defined over my hotdogs and my hamburgers. But, of course, 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 (homo economicus) or other-regarding (including preferences for equity). As such, the received results of giving in *standard* dictator games, while inconsistent with homo economicus, can be accommodated by convex preference theory (Andreoni and Miller, 2002; Fisman et al. 2007).

2.2 Experiments in which Behavior is Inconsistent with Convex Preference Theory

List (2007) and Bardsley (2008), amongst others, use 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 now discuss.

Figure 1 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[7] – a pattern that the data clearly refute. Convex preference theory also implies that the choices that are in the interior of the

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

[7] Indifference curves representing dictator's preferences for own payoff and recipient's payoff that are strictly convex to the origin imply that anyone who chooses 0 or -1 in the Take 1 treatment will choose 0 in the Baseline.

feasible sets for both the Baseline and Take 1 treatments should be the same. The data are also inconsistent with this prediction of convexity. 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 implications as conventional 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 these dictator game data.

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

Convexity, however, is not necessary for choice rationality, so comparisons such as the above for the List and Bardsley data are uninformative about choice rationality. An illustration of rational choices for non-convex preferences is shown in Figure 2. An individual with such preferences will choose x in the give or take game with the feasible set [A, C] and the endowment at point B. The same individual, however, will choose y (rather than B) in the give game with feasible set [A, B] and endowment at point B.

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

Convexity and rationality 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). With *C(S)* and *C(T)* denoting the choice sets when the feasible sets are S and T the Contraction Consistency Axiom (CCA, also known as Property α from Sen 1971, 1986) states: For all feasible sets S and 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. For single-valued choice functions, CCA is a necessary and sufficient condition for existence of a complete and transitive order (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.

A specific implication of rational choice theory is for each of the bars portraying fractions of choices of $0 to $5 in the Take 1 treatment to be no 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 though not with convexity. Similarly, data from Experiment 2 conducted by Bardsley (2008) refute convexity but are mostly consistent with rational choice theory; the bar at $1.50 (two observations) is the only inconsistency with rationality in Experiment 2.

2.3 Experiment in which Behavior is Inconsistent with Contraction Consistency Axiom

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 requires choices be invariant to changes in the endowments in the experiment: for any two endowments, the choice set remains the same set. Let S1 ($4.05) denote the average payoff of $4.05 to the recipient in scenario 1. Using this same notation to reflect payoffs in the remaining scenarios, 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 endowment treatments is inconsistent with predictions from CCA.

3. MORAL MONOTONICITY THEORY

The systemic 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) postulation 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 introduce 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

The items of choice are n-vectors of amounts of money (or some other defined "good"). 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 $\boldsymbol{f}^*$ be chosen from some feasible set F and let G be a subset of F that contains it.[8] Let $\boldsymbol{r}^G$ and $\boldsymbol{r}^F$ denote moral reference points (to be defined below) for feasible sets G and F . If the moral reference point in G is more favorable to individual $i,$ and (weakly) less favorable to others, then we postulate that no choice from G allocates individual i less than $\boldsymbol{f}^*$. Similarly, if the moral reference point in G is less favorable to individual i, and (weakly) more favorable to others, then no choice from G allocates individual i

[8] Bold font is used to denote vectors.

more than $\boldsymbol{f}^*$. Formalizing this, for every individual $i\,(=1,\cdots,n)$ one has:

MORAL MONOTONICITY AXIOM (MMA):

a. If $G \subseteq F$, $r_i^G \geq r_i^F$ and $\boldsymbol{r}_{-i}^G \leq \boldsymbol{r}_{-i}^F$, then $\boldsymbol{f}^* \in F^* \cap G \Rightarrow g_i^* \geq f_i^*, \forall \boldsymbol{g}^* \in G^*$

b. If $G \subseteq F$, $r_i^G \leq r_i^F$ and $\boldsymbol{r}_{-i}^G \geq \boldsymbol{r}_{-i}^F$, then $\boldsymbol{f}^* \in F^* \cap G \Rightarrow g_i^* \leq f_i^*, \forall \boldsymbol{g}^* \in G^*$

where F* and G* are choice sets of F and 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[9] 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[10]

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

For singleton choice sets, this requires $\boldsymbol{f}^*$ to be *the* chosen allocation in any subset, G of F that contains $\boldsymbol{f}^*$. Implications of MMA for choices is stated in the following proposition.[11]

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 Korenok et al. (2014) experiment and (b) the List (2007) and Bardsley (2008) experiments. All of these experiments include taking as well as giving opportunities. The Korenok et al. experiment varies the dictator's endowment while holding constant the minimum (at other's maximum) payoff of each agent at 0. Their data suggest that dependence of choices on dictator's endowment in a way not captured by traditional theory is

---

[9] Recall that one of the attractiveness of (strict) convex preferences is that choice sets are singletons.

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

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

empirically significant.[12] 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 (at other's maximum) payoffs. Taken together, these experiments suggest that choice behavior is dependent on (a) the dictator's endowment and (b) the minimum (at other's maximum) payoffs available in the game. We define moral reference points inspired by this experimental literature.

Our definition of moral reference point incorporates two intuitions into the 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, x that benefits me at your expense decreases with the closeness of my final payoff to my endowed ("status quo") payoff: my "property right." Intuition (b) reflects the idea that my moral cost from making such a choice x:

(i) *decreases* with the (positive) difference between your final payoff (x_2) 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 (x_1) 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 insights 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 MMA 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 and m^o and y^o be the maximum feasible payoffs for the dictator and the recipient in F, that is

[12] 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 recipient's endowment. Assumed dependence on *both* endowments would be an over-determined miss-specification because they sum to a constant.

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

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, that is

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

Following intuitions (a) and (b), 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,

$$(*) \qquad \boldsymbol{r}^F = ((1-\theta)m_*(F) + \theta e_m, y_*(F)),$$

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).[13] In all our applications the endowments are assigned. So, to provide a simple, transparent way of identifying moral reference points throughout the paper, we use $\theta = 1/2$. However, all of the analysis in this paper can be easily replicated with any other weight $\theta \in (0,1)$ and such changes would have no impact on the comparative statics underlying the empirical analysis in Sections 5 and 6.

An illustration on how to locate moral reference points is shown is Figure 4, using the example $\theta = 1/2$, and Give, Take, and Symmetric action sets for treatments in the students experiment (reported in Section 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 these steps in Figure 4 using

[13] The intuition that "property rights" matter for final allocations in a dictator game is consistent with results in Cherry et al (2002), Oxoby and Spraggon (2008) and Korenok et al. (2017). Cherry et al (2002) report that dictators share less with the recipient when dictators earned the endowment. Oxoby and Spraggon (2008) and Korenok et al. (2017) 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. If $\theta=1/2$ when endowments are assigned, then $\theta>1/2$ for dictators earning the endowment and $\theta<1/2$ for responders earning the total endowment.

three dictator games as examples. When, in the Symmetric action set, the endowment is at point *B* and feasible set contains discrete points on the budget line extending from *A* to *C*, the minimal expectations point is the lower left corner of the large triangle and the moral reference point is r_S. For the Give action set with endowment at point *C* and the budget line extending from *B* to *C*, the minimal expectations point is the lower left corner of the small triangle and the moral reference point is r_G. Finally, for the Take action set and endowment at point *B*, the budget line extends from *B* to *C* and the minimal expectation and moral reference point are both at r_T.

[FIGURE 4 ABOUT HERE: Examples of Moral Reference Points]

3.3 MMA and WARP

In some contexts, testable implications of MMA are the same as the weak axiom of revealed preference (WARP). Consider, for example, the dictator game experiment of Andreoni and Miller (2002) that varied underlying budget sets and applied the generalized axiom of revealed preference (GARP) to analyze the consistency of choices.

Figure 5 illustrates two budget sets like those that the dictator can face in the Andreoni and Miller design. Let point *a* (on the horizontal axis) denote the initial endowment on the steeper line and point *b* denote the initial endowment on the flatter line. Further, consider the shaded quadrilateral that is the intersection of the opportunity sets bounded by the steeper and flatter budget lines. Viewed through the lens of MMA, let the shaded quadrilateral set be a feasible set with endowment at point *a*. The minimal expectations point is the origin (0,0) for all three feasible sets. 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 5 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. The quadrilateral set is a subset that contains A, and has the same moral reference point (r^a), so by MMA (see Proposition 1), A is also chosen from the quadrilateral set. The main implication of MMA here is identical to CCA. Next, 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 and contains B) allocates to the dictator less than B does, because r^a is to the left of r^b. Choice of A from the quadrilateral set violates this. Thus, any pair of choices of type A and B here that violate WARP also violate MMA..

3.4 MMA and Data from Some Dictator Games in the Literature

CCA and MMA predictions for behavior in the Korenok et al. (2014) experiment are different. In all of their treatments, the minimal 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 statement (∗) for their several endowment treatments are shown in Figure 3 using $\theta = ½$ for ease of illustration as endowments were assigned. 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. MMA requires dictator's choices to decrease the amount allocated to oneself for each change in endowment from 1 to 9, while CCA or convex preferences require the choice remain the same. Korenok et al.'s (2014) data reject CCA in favor of MMA in three out of four comparisons (and the change from endowment 8 to 9 is insignificant).

Turning attention back to the experiments reported by List (2007) and Bardsley (2008), we note that while their data refute convexity they are consistent with CCA and MMA. Their experimental designs, however, have little power for testing either CCA or MMA because the contraction (baseline) does not contain choices of more than 40% of their subjects (who took in the Take treatments) We next explain two new experiments we designed to test MMA directly.

4. EXPERIMENTAL DESIGNS AND PROTOCOLS

Following List (2007), our designs begin 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 feasible set of final allocations. The crossed design, of varying the action sets as well as the endowments, exogenously varies moral reference points, which allow us to identify the importance of such on observed choices.[14]

4.1 Design for the Children Experiment

Figure 6 shows the three major treatment types in the experiment with children denoted "Inequality," "Equal," and "Envy." The Inequality-Give treatment represents a typical dictator game: both the dictator and recipient have a fixed endowment of 4 units, which corresponds to the "show-up fee" in comparable laboratory experiments. In addition, the dictator is provided a variable endowment of 4 units (so the endowment is at point B_I) and can choose to give all, some, or none of that amount to the recipient. In the Inequality-Take treatment, the fixed endowment is the same as in the Inequality-Give treatment, but now the "property rights" for the 4 unit variable endowment are assigned to the recipient rather than the dictator (so the endowment is at point A_I) but the dictator can take none, some, or all of that amount. In the Inequality-Symmetric treatment, the endowment is at point B_I and we expand the action set by allowing the dictator to either give any amount from 0 to 4 or take any amount from 0 to 4. Across all treatments, we restrict the choices of the dictator such that only integer amounts can be given or taken.

[FIGURE 6 ABOUT HERE: Feasible Sets for the Children Experiment]

The only difference between the Inequality treatments and the Equal and Envy treatments is that we vary the manner in which the fixed endowment is distributed across the dictator and the recipient; this shifts the feasible sets northwest and preserves the price of giving. Specifically, in the Equal treatment, we move 2 from the dictator's fixed endowment into the recipient's fixed endowment. In the Envy treatment, we move 4 from the dictator's fixed endowment into the recipient's fixed endowment.

In both the Inequality-Give and the Inequality-Symmetric treatments, the initial endowment of (8,4) favors the dictator and she is thus faced with an allocation decision over a

[14] Our approach to identifying the importance of moral reference points shares similarity with Krupka and Weber (2013) who test the importance of norms by comparing final allocations across a standard dictator game and what they call the Bully treatment where the initial endowment is split amongst the dictator and recipient and the dictator is allowed to either give to or take from the recipient. Kimbrough and Vostroknutov (2015) use an alternative approach to identify the importance of norms on dictator behavior by eliciting individual-specific measures of norm-sensitivity and correlating this with observed allocations in the standard dictator game.

budget set that crosses the 45-degree line, as in most standard dictator games: see Figure 6. In the Equal-Give and Equal-Symmetric treatments, the initial endowment of (6,6) lies on the 45-degree line. However, the treatments differ in that the feasible budget set for the Equal-Give treatment lies on and above the 45 degree line, whereas the feasible budget set for the Equal-Symmetric treatment crosses the 45 degree line.[15] In the Envy-Give and the Envy-Symmetric treatments, the initial endowment (4,8) lies strictly above the 45-degree line and favors the recipient. Yet, these treatments differ as the feasible budget set for the Envy-Give treatment lies strictly above the 45-degree line, whereas the feasible budget set for the Envy-Symmetric treatment again crosses this line.[16]

The nine treatments of the children design provide a test of the central properties of MMA: monotonicity of choice in each dimension of moral reference point. Table 1, first column, shows the moral reference points (in parentheses) for the nine treatments in the children experiment.

4.2 Protocol for the Children Experiment

The experiment was conducted at the Chicago Heights Early Childhood Center. Children were either brought in by parents at a designated time outside normal school hours, or participated during school hours by being taken out of class. All children who participated were assigned the role of the dictator, while children whose parents did not bring them in for the study played the role of receivers and were sent their final payoff via the mail.

Each child participated in only one session of the experiment, during which he/she was assigned to either the Inequality, Equal, or Envy treatment. Preschool-age children (ages 3 to 4) are predominant in our sample and were randomly assigned to one of the three treatments with equal probability. Kindergarten and 1st grade-age children (ages 5 to 7) were always assigned to the Inequality treatment. Children encountered Give, Symmetric, or Take action sets in random order. After the first decision was paid out, children were surprised with two additional dictator games with the remaining two action sets (which were also paid out).

[15] In this regard, our design shares a similarity with treatment 1 in Cox and Sadiraj (2012) which has dictators make allocations over a budget set that lies on and above the 45-degree line as a means to test a defining characteristic of models of inequality aversion.

[16] Note that "inequality," "equality" and "envy" do not apply literally in the Take treatments, since in the Take treatments the initial endowment always favors the recipient as property rights to the variable endowment are assigned to the recipient. Yet in all such cases, the budget set over which the dictator is selecting a final payoff is identical to that faced in the corresponding Give treatment.

Following prior experiments with very young children (e.g., Li et al., 2013), we used stickers as the payoff medium. To further ensure saliency and dominance of payoffs, we first gave the child an option to select one of two predetermined sticker sets as the payoff medium in each dictator game.[17] In order to conduct the experiment with children, we designed specialized receptacles, as shown in Figure 7. One receptacle belonged to the dictator and the other to the intended recipient. In each treatment, both the dictator and recipient started out with a number of stickers that could not be moved (the fixed endowment). These stickers were housed inside of clear boxes. The stickers available for distribution (the variable endowment) were displayed on plates that were on top of the clear boxes.

[FIGURE 7 ABOUT HERE: Experiment Setup for the Children Experiment]

Children could move stickers from plate to plate until they were satisfied with their choice. Once the decisions were made, the stickers remaining on each plate were moved into the corresponding boxes. At the beginning of the experiment, instructions were read aloud by the experimenter to explain how these boxes and plates were to be used. The instructions also included questions to ensure that children understood the game. Children received a small toy at the end of the experiment as a "show up fee." The toy was pre-announced in the recruitment letter, but experimenters did not remind children of the toy prior to the start of the experiment. Parents who brought children to the session received $10 for their time and an additional $5 for completing a short survey while they waited. Each session lasted approximately 10-12 minutes.

4.3 Participation in the Children experiment

In total, we had 329 children participate in our first experiment. The average age of our subjects was 5, with the majority of children (183) below the age of 5; the minimum age was 3.5 and maximum age was 7.4 years old. In our sample, we have 50% males, 45% Hispanic, 43% African-American, and 12% Caucasian.[18] Treatments are balanced on demographics, with the exception of the Inequality treatment, which is unbalanced on age by design, as aforementioned.[19]

---

[17] Available sticker types were star or mustache, underwater or space, and cat or dog, respectively, in the first, second, and third dictator games.

[18] We do not have a race on file for two subjects. Age is missing for one subject.

[19] Conducting regressions (that include experimenter fixed effects) of all treatment dummies with either gender or race (separately for African-American, Hispanic, or White) does not yield any significant coefficients for

4.4 Design for the Undergraduate Student Experiment

We also ran an experiment with subjects recruited from the common convenience sample of undergraduate students. This experiment also shifted the action sets and endowments with the aim of increasing the proportion of observations in the intersection of the Symmetric action set with the Give and Take action sets, which would provide increased opportunity to observe inconsistencies with CCA and MMA.

Figure 8 shows three budget lines for the students experiment labeled "Inequality," "Equal," and "Envy." The finite feasible sets include discrete points on the lines. Labeling 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 8 ABOUT HERE: Feasible Sets for the Students Experiment]

The experimental design is 3 $\times$ 3: (Inequality, Equal, Envy) $\times$ (Symmetric, Take, Give). The sum of the payoffs of dictator and recipient is the same (30) in all nine treatment cells. In the Inequality-Give treatment (with endowment at point C_I in the left panel of Figure 8): the 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

demographic characteristics. Conducting similar regressions with age at time of the test as the dependent variable for each budget line treatment (Inequality, Equal, or Envy) does not yield significant differences by action set.

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.[20]

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 of our students design also provide a direct test of MMA: monotonicity of choice in both dimensions of moral reference point. Table 1, third column from the right, shows the moral reference points (in parentheses) for all nine treatments in this experiment.

4.5 Protocol for the Undergraduate Experiment

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 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.

[20] 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.

Subjects were informed that there was no show-up (or non-salient participation) fee in this experiment.[21] 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.

4.6 Participation in the Undergraduate Experiment

In total, we had 612 subjects (306 dictators) participate in the undergraduate student 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).

5. EXPERIMENTAL RESULTS

5.1 Overview of Results

Table 1 reports summary statistics for both experiments. The left-most column reports the feasible actions (Symmetric, Give, or Take) and the moral reference points for the Inequality, Equal, and Envy endowment treatments in the children experiment. The data are largely in line with our theory.

Beginning with the Inequality endowment treatment in the children experiment, we observe that the moral reference points change from (6,0) to (6,4) to (4,4) as the action set is changed from Symmetric to Give to Take. Negative monotonicity with respect to the recipient's

---

[21] The purpose of this design feature was to avoid the ambiguity that would follow from an experimenter not knowing whether a dictator integrates or does not integrate positive show up fees with salient payoffs in deciding how much to give or take. The feasible sets were constructed so that no subject in any treatment could leave the lab with less than \$3.

moral reference dimension, r_2 implies that the average dictator payoff will be lower with the Give action than with the Symmetric action space. The average payoff of 6.67 is lower than 8.79, as reported in the second-left column (with standard deviations in brackets). Positive monotonicity with respect to the dictator's moral reference dimension, r_1 implies that the average dictator payoff will be lower with the Take action than with the Give action space. We observe exactly that result. Furthermore, the relative sizes of the average dictator payoffs in the other six rows of the first column of Table 1 are as predicted by MMA.

The fourth column of Table 1 reports the moral reference points for all treatment cells in the students experiment. As predicted by MMA, the observed average dictators' payoffs in Give are larger than in the Symmetric action sets for all three treatments (Inequality, Equal and Envy). The differences between observed relative values of average dictators' payoffs in Give and Take are not as predicted by MMA for Inequality and Envy treatments.

5.2 Analysis of Data from Children Experiment

To evaluate whether moral reference points affect allocations as predicted under MMA, we pool data from the three rounds of the children experiment to estimate two Tobit models – one that conditions choice solely upon moral reference point, r and a second that augments this specification to include dictator fixed effects – and estimate the models separately for three moral reference points in the experiment. To control for potentially binding budget constraints, we set bounds to establish a common support across games with a given r. The lower (upper) bound in each model is the lowest (highest) possible payoff a dictator could receive in the common support.

Under standard models of choice, the moral reference point should have no influence on payoffs. Hence, we would expect the estimated coefficients on the indicators for the r_1 and r_2 dimensions of the moral reference point to be zero. In contrast, payoffs depend on the moral reference point under MMA. Specifically, MMA predicts that dictator payoffs are (i) decreasing in r_2 and (ii) increasing in r_1. We would thus expect a positive coefficient on the r_2 dimension of the moral reference point as the indicator variable in our model captures choices where r_2 is less than a baseline level r_2^c. Similarly, we would expect a negative coefficient on the r_1 dimension of the moral reference point as the indicator variable in our model captures choices where r_1 is less than a baseline level r_1^c.

Table 2 reports Tobit estimated coefficients for dummy variables for r_1 and r_2 dimensions of moral reference point using the pooled, within-subjects data for the children experiment. Results are reported with and without child fixed effects. The first two columns of the table restrict the analysis to choices from the Inequality treatment and use $r = (6,4)$ as the baseline reference point. The third and fourth columns restrict the analysis to choices from the Equal treatment and use $r = (4,6)$ as the baseline reference point. The final two columns correspond to data from the Envy treatment and baseline moral reference point of $r = (2,8)$.

TABLE 2 ABOUT HERE: TESTS FOR EFFECTS OF MORAL REFERENCE POINTS

Equal Treatments: We begin by analyzing data from the Equal treatment. Every child in this treatment made choices for each possible action – Give, Take, and Symmetric – with the order of these actions randomized across subjects. The feasible payoff for the dictator in both the Equal-Give and Equal-Take treatments is an integer (weakly) between 2 and 6. The feasible payoff for the dictator in the Equal-Symmetric treatment is an integer (weakly) between 2 and 10. Hence, the budget sets for Equal-Give and Equal-Take are contractions of the budget set for Equal-Symmetric but all three contain integers (weakly) between 2 and 6.

As noted in Table 1, the average payoff for the dictator across these actions are (i) 5.03 in Equal-Give, (ii) 4.50 in Equal-Take, and (iii) 6.84 in Equal-Symmetric. To control for differences in the budget constraint (feasible payoffs) and test whether or not observed payoffs are influenced by moral reference points as predicted by MMA, we run Tobit regressions of dictator's final payoffs that use lower and upper bounds of 2 and 6. To allow independent variation in each dimension of the moral reference point, we use data from the Equal-Give treatment as our baseline and include indicators for the other two treatments.[22]

Results from the Tobit regression are included in Table 2 and provide evidence consistent with MMA and the importance of *both* the r_1 and r_2 dimensions of the moral reference point on dictator payoffs. For example, the moral reference point is less favorable to the dictator in the

[22] The moral reference point is r = (4, 6) in the Equal-Give treatment. The moral reference point in the Equal-Take treatment, r = (2, 6), differs only in the r_1 dimension and the moral reference point in the Equal-Symmetric treatment, r = (4, 2) differs only in the r_2 dimension. Hence, the indicator variables for observations where $r_1 < r_1^c$ and those where $r_2 < r_2^c$ are equivalent to indicators for the Equal-Take and Equal-Symmetric treatments, respectively.

Equal-Take treatment than it is in the Equal-Give treatment. Since the feasible sets in both treatments are identical, the standard model of choice would predict no difference in dictator payoffs. MMA, in contrast, predicts that dictator payoffs should be lower in the Equal-Take treatment. This is precisely what we observe in our data. As noted in the middle columns of the table, the estimated coefficient on the indicator variable for observations where $r_1 < r_1^c$ is negative and statistically significant at the $p < 0.05$ level both with and without child fixed effects.

We observe similar evidence when exploring changes in the r_2 dimension of the moral reference point. The moral reference point is less favorable to the recipient in the Equal-Symmetric treatment than in the Equal-Give treatment. MMA thus predicts that the recipient's payoff in the Equal-Symmetric treatment should be lower in Equal-Symmetric than in Equal-Give. As total payoffs across treatments are held constant at 12, this implies that the dictator's payoff in Equal-Symmetric should be greater than those observed in Equal-Give, i.e., the estimated coefficient on the indicator variable for observations for the Equal-Symmetric treatment should be positive. This is again what we observe. As reported in the middle columns of the table, the estimated coefficient for observations where $r_2 < r_2^c$ is positive and statistically significant at the $p < 0.01$ level, both with and without child fixed effects, suggesting that dictators keep more for themselves in the Equal-Symmetric treatment.[23]

Inequality and Envy Treatments: We follow the same approach when analyzing data from the Inequality and Envy Treatments. We observe three choices for subjects in these treatments – one each from the Give, Take, and Symmetric action sets – and use data on dictator payoffs to estimate a Tobit model with bounds set by the common support for dictator payoffs in the corresponding treatment.[24] As before, we estimate two variants of the model for each treatment – one with individual fixed effects and another without such fixed effects – and include indicators for observations from the respective Take ($r_1 < r_1^c$) and Symmetric ($r_2 < r_2^c$) treatments.

Data from these treatments provide further evidence consistent with the predictions of MMA, and are at odds with the conventional model of choice. Specifically, we find that changes

[23] As is evident from the standard errors, all of the $r_2 < r_2^c$ test results in Table 2 reported as significant at p<0.01 are also significant at p<0.001 except the entry in the right-most column which is significant at p<0.002.

[24] The common support for dictator payoffs across the three action spaces is (4, 8) in the Inequality treatment. We thus set 4 as a lower bound in these models and 8 as the upper bound. The common support for dictator payoffs in the Envy treatment is integers in [0,4] and serve to define the lower and upper bounds for the Tobit.

in both dimensions of the moral reference point influence dictator payoffs. For example, the estimated coefficient on the indicator for observations in the Take treatment – i.e., observations where $r_1 < r_1^c$ – is negative and statistically significant at the p < 0.01 level for the Envy treatment; results for the Inequality treatment are similar but less significant. The estimated coefficients for observations in the Symmetric treatment – i.e., observations where $r_2 < r_2^c$ – are positive and statistically significant at the p < 0.01 level for both the Inequality and Envy treatments with and without child fixed effects.

Viewed in its totality, data from our children experiment are consistent with the predictions of MMA and highlight the importance of moral reference points. This result is buttressed by the fact that the sign on every coefficient estimate in Table 2 is consistent with the predicted effect under MMA.

5.3 Analysis of Data from Undergraduate Student Experiment

As a second test of our theory, we rely upon a between subjects comparison utilizing data from the undergraduate student experiment. Recall that MMA predicts that the recipient's payoff increases in r_2 when r_1 is fixed. Since the total available payoff in our experiment is fixed, comparative statics that predict an increase in the recipient's payoff necessarily imply a reduction in the payoff for the dictator and vice versa. In contrast, conventional choice theory, and any model of convex preferences, predict that changes in r_2 should have no impact on final payoffs.

To test for r_2 effects, we use a between-subjects comparison across three different values of r_1.[25] By design, the r_1 dimension of the moral reference point within a given treatment (Inequality, Equal, or Envy) differs across the possible actions (Give, Take, Symmetric). Hence, testing the effects of changing r_2 while holding r_1 constant requires that we compare dictator payoffs across treatments. Since all subjects were randomly assigned to a single treatment, this means that the corresponding tests will compare payoffs across individuals – a between subject analysis.

---

[25] There are five possible values of r_1 in our experiment: 7, 11, 15, 19, 23. There is only one treatment (Envy-Symmetric) with r_1 = 7 and only one treatment (Inequality-Give) with r_1 = 23. As there is no variation of r_2 with these two r_1 values we can't use data from these two treatments to directly test MMA in terms of predicted r_2 effects.

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) – for each of the three levels of r_1 in our experiment. 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 each model, we define a baseline moral reference point (r_{1i}^c, r_{2i}^c) and include indicator variables that equal one for observations where either (i) $r_{2i} > r_{2i}^c$ or (ii) $r_{2i} < r_{2i}^c$. This differs from our analysis of the children data where we only observed choices for which $r_2 > r_2^c$. Under standard models, the estimated coefficient on both indicators should be zero. Under MMA, the estimated coefficient on the indicator for $r_{2i} < r_{2i}^c$ should be positive and the coefficient on the indicator for $r_{2i} > r_{2i}^c$ should be negative. Table 3 presents results for these models. The left two columns restrict the analysis to the subset of data where r_1 = 15. The middle two columns restrict the analysis to data where r_1 = 19 and the right two columns to data where r_1 = 11.

TABLE 3 HERE: TESTS FOR EFFECTS OF RECIPIENT MORAL REFERENCE DIMENSION

Data from $r_1 = 15$: There are three different treatments where the dictator's dimension of the moral reference point (r_1) is held constant at 15: Inequality-Symmetric (r_2 = 3), Equal-Take (r_2 = 7) and Envy-Give (r_2 = 11). As noted in Table 1, average payoff for the dictator across these three treatments is decreasing in the level of r_2: 20.88 in Inequality-Symmetric, 19.83 in Equal-Take, and 16.57 in Envy-Give. While such patterns are consistent with predictions of MMA, each treatment has a distinct set of feasible payoffs, which confounds tests based on raw averages.[26] However, as each treatment has a common support [15, 19], we can use these as bounds in a Tobit model and thus control for differences in budget constraints across treatments.

As we observe three distinct values of r_2, we can estimate the effects of both increases and decreases in the recipient's dimension of the moral reference point. To do so, we set $r_2 = 7$ and include indicator variables for observations from the Inequality-Symmetric ($r_2 < r_2^c$) and Envy-Give ($r_2 > r_2^c$) treatments. Empirical results are presented in the first two columns of Table 3 and provide mixed support for the predictions of MMA. Consistent with the predictions of MMA, the estimated coefficient on the indicator for the Envy-Give treatment is negative and statistically significant at the $p < 0.01$ level. This suggests that as the moral reference point becomes more favorable to the recipient, dictators keep less for themselves. The estimated coefficient on the indicator for the Inequality-Symmetric treatment ($r_2 < r_2^c$) is of the wrong sign. As the moral reference point in this treatment is less favorable to the recipient, MMA predicts that dictator earnings should be larger than in the baseline treatment. The estimated coefficient on this indicator is negative, but the estimate is imprecise and not statistically significant at any meaningful level.

Other Reference Points: We follow the same approach to analyze data for choices where the dictator's dimension of the moral reference point takes other values – i.e., choices where $r_1 = 19$ and $r_1 = 11$. To estimate the effect of changing r_2, we use data on dictator payoffs to estimate a Tobit model with bounds set by the common support for dictator payoffs in the corresponding treatment.[27] As before, we estimate two variants of the model for each treatment – one with demographics and another without such controls – and include an indicator variable for choices where $r_2 > r_2^c$. For choices where $r_1 = 19$ this corresponds to estimating an indicator variable for choices from the Equal-Give treatment and for choices where $r_1 = 11$ this corresponds to estimating an indicator variable for choices from the Envy-Take treatment.

[26] The set of feasible payoffs across the three treatments are: (i) [11, 27] in Inequality-Symmetric, (ii) [15, 23] in Equal-Take, and [14, 19] in Envy-Give. Note that feasible (budget) sets for Equal-Take and Envy-Give are contractions of the feasible set for Inequality-Symmetric.

[27] The common support for dictator payoffs across the two treatments where $r_1 = 19$ is integers in [19, 23]. We thus set 19 as a lower bound in these models and 23 as the upper bound. The common support for dictator payoffs in treatments where $r_1 = 11$ is integers in [11, 19] and serve to define the lower and upper bounds for the Tobit.

Data from these models provide further evidence consistent with the predictions of MMA, and at odds with the conventional model of choice. For example, as reported in the middle columns of Table 3, the effect of increasing the recipient's dimension of the moral reference point is a reduction in dictator payoffs of approximately \$1.56 to \$1.66 – differences that are statistically significant at the $p < 0.10$ level.[28] Similarly, the estimated coefficient for observations in the Envy-Take treatment – i.e., observations where $r_2 > r_2^c$ given that $r_1 = 11$ – is negative and statistically significant at the $p < 0.05$ level.

Viewed in their totality, results from the undergraduate student experiment are largely consistent with predictions of MMA, and hold when we include demographic controls. Specifically, we find statistical support for three of the four tests of r_2 effects. Interestingly, in all such instances, our results suggest that dictators keep less for themselves as the moral reference point becomes more favorable to the recipient. We view this as strong evidence that moral costs matter, and a result that is consistent with prior work showing that individuals are willing to forego potential benefits to avoid social pressures or guilt that arise when acting "selfishly" (e.g., Della Vigna et al., 2012; Ferraro and Price, 2013).

5.4 Alternative Models

We briefly examine implications of alternative models of behavior: selfish preferences, social preferences, reference dependence, and sharing and sorting.

Selfish Preferences: Two-thirds of the transfers are positive and four out of five of our dictators made at least one positive transfer. Given our restriction of data analysis to common supports across treatments, the selfish preferences model predicts changes in r_1 or r_2 will have no effect. Parameter estimates for both are statistically significant, rejecting the implications of 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.

[28] As is evident from the standard errors, the test results in Table 3 reported as significant at $p<0.01$ are also significant at $p<0.001$.

Reference Dependent Models. The status quo (initial endowment) is the reference point in the classical loss-aversion reference dependent model of Tversky and Kahneman (1991). This TK model predicts that the dictator's final payoff allocation in the Give treatment is larger than in the Take treatment, which is the same as our MMA prediction. This is so because in the Give scenario all feasible allocations introduce loss on dictator's dimension and gain on recipient's dimension whereas in the Take scenario, all feasible allocations offer gain for dictator's payoff but loss for recipient's payoff. However, in the Symmetric and Take scenarios (in the undergraduate students experiment) the status quo (the initial endowment) is the same, and therefore the prediction of this model is the same as the CCA prediction. MMA predicts a larger final allocation for the dictator in the Take scenario when choice in Symmetric is between B and C. For the children experiment, TK model's prediction is the same as CCA's for the Symmetric and Give scenarios, whereas MMA predicts a smaller final allocation for the dictator in the Give scenario when choice in Symmetric is between A and B. In summary, the data support MMA, not TK.

Koszegi and Rabin (2006) model of reference dependence has recently seen a surge in applied work. Predictions of this KR 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 KR model.[29] Because our data reject conventional theory, the KR model is also rejected.

Sharing and Sorting. Lazear et al. (2012) offer a model of sharing that depends on the environment, where an indicator variable 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.

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 provide compelling evidence that observable moral reference points matter, and influence choice in a manner consistent with MMA.

6. IMPLICATIONS OF MMA FOR OTHER TYPES OF DICTATOR GAMES

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

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 experiments. Yet, MMA has broader implications for choice in a range of related experiments including, as explained above, standard (give-only) dictator games (Andreoni and Miller 2002), and other dictator games that compare the effects 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 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 predictions of the conventional rational choice model. We have explained above that their data are consistent with MMA.[30]

6.2 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).

[30] 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 data analysis in Dreber et al. (2013) relies upon a normalized metric of taking or giving that codes transfers in the Take only treatment as positive instead of negative; hence the seeming inconsistency with MMA is not actually present in the data. 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 potentially 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.

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.

6.3 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.[31] The Lazear et al. Experiment 2 is a within-subjects design including several decisions with one selected randomly for payoff. In Decision 1, 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.[32]

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 monotonicity model, in which

[31] 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)".

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

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 $(r_1, r_2) = (\frac{1}{2} \times 10, 0)$. An example of choices consistent with MMA, can be captured by maximization of $u(m - r_1, y - r_2)$ for some increasing function $u(\cdot)$. 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 (see Appendix C for an example).

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 et al. (1996).[33] 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 any $z > 0$.

6.4 MMA and Earned Endowments

[33] The exact wording in the Hoffmann et al. (1994) 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)".

Oxoby and Spraggon (2008) report an experiment with dictator games that includes treatments in which 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 (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 data 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 the 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 (or GARP).

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.[34] 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

[34] 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.

"prediction" results solely from the constraint that prevents taking in the investment game, not from implications of CCA itself.

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, then 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 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

$$\begin{aligned} M_s^+ &= \{(e-s+r, e+ks-r) : r \in [0, ks]\} \\ M_s^- &= \{(e-s+r, e+ks+r/k) : r \in [-(e-s), 0)\} \end{aligned}$$

Consider the SM's choice in M_s in the moonlighting game when the FM sends a non-negative amount. Consistent with observed behavior[35] (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,

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

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.

[35] In data reported by Cox, Sadiraj, and Sadiraj (2008) Only 2 out of 46 second movers who did not have money taken from them by first movers chose $r_s \in M_s^-$.

<u>FM choices across the two games</u>: 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

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

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 .

<u>Implications for game play</u>: 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 choose a larger amount 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.

<u>Existing data that provide empirical support for MMA</u>: 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.

<u>FM choices</u>: 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.[36] Therefore, the FM data are consistent with the above implications of MMA but inconsistent with implications of CCA.

<u>SM choices</u>: Estimates (standard errors in parentheses) of censored regressions for SM choices at information sets with "FM not taking" (send ≥ 0 , N=78) are[37]

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

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) explore the 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)

---

[36] If we examine only at Send > 0, averages are 7.35 (IG) and 4.84 (MG), which are significantly different (t-test, p-value=0.004) at conventional levels.

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

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 $5r_s$.

<u>*SM choices across the three games*</u>: 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 that $M_{cs}^s = M_c^s \cup M_s^s$. 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.

In sum, received 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, this 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.

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 choice theory to increase its empirical validity. The Korenok et al. (2014) work suggests that choices depend on endowments in ways not captured by conventional theory. The List (2007) and Bardsley (2008) studies 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 on two novel experiments designed to test the central feature of the new theory: monotonicity in choice with respect to distinct dimensions of observable moral reference points. Data from the experiments largely reject CCA in favor of MMA.

The theory of moral monotonicity, 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 cooperative 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.

Cherry, Todd L., Peter Frykblom, and Jason F. Shogren. “Hardnose the Dictator.” *American Economic Review* 92, no. 4 (2002): 1218-1221.

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. “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.

Li, Yiyuan, Hong Li, Jean Decety, and Kang Lee. "Experiencing a Natural Disaster alters Children's Altruistic Giving." *Psychological Science* 24, no. 9 (2013): 1686-1695.

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.

Tversky, Amos, and Daniel Kahneman. “Loss aversion in Riskless Choice: A Reference-Dependent Model.” *The Quarterly Journal of Economics* 106, no.4 (1991): 1039-1061

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

	Children Experiment			Undergraduate Experiment		
	Action[a]: Moral Ref.	**Dictator Payoff**	**Nobs**	**Action[a]: Moral Ref.**	**Dictator Payoff**	**Nobs**
Inequality	S: (6, 0)	8.79 {1.77}	163[b]	G: (23, 3)	22.46 {2.96}	61
	G: (6, 4)	6.67 {1.10}	164	T: (19, 3)	22.81 {3.34}	81
	T: (4, 4)	6.42 {1.18}	164	S: (15, 3)	20.88 {4.95}	82
Equal	S: (4, 2)	6.84 {1.61}	88	G: (19, 7)	20.35 {2.24}	66
	G: (4, 6)	5.03 {1.20}	88	T: (15, 7)	19.83 {2.88}	58
	T: (2, 6)	4.50 {1.37}	88	S: (11, 7)	19.06 {3.52}	62
Envy	S: (2, 4)	5.29 {2.04}	77	G: (15, 11)	16.57 {1.65}	67
	G: (2, 8)	3.40 {1.09}	77	T: (11, 11)	16.94 {1.85}	69
	T: (0, 8)	2.82 {1.41}	77	S: (7, 11)	16.36 {2.55}	66

[a] Notation for Action: S (Symmetric), G (Give), T(Take); [b] In the Inequality treatment, one child made only two decisions, Give and Take treatments only. Moral Reference Points in parentheses with dictator dimension first followed by recipient dimension; Standard deviations in brackets; Nobs is the number of observations in each treatment (children made three decisions, students made two decisions); In all treatments, the sum of dictator and recipient payoff is 12 in Children Experiment and 30 in Students Experiment.

Table 2. **Effects of Moral Reference Point (Children experiment; within subjects)**

Dep. Var: Dictator Payoff	Inequality $r^c = (6,4)$		Equal $r^c = (4,6)$		Envy $r^c = (2,8)$	
$r_1 < r_1^c$ [-]	-0.345	-0.361**	-0.992**	-1.058***	-1.867***	-1.854***
	(0.215)	(0.153)	(0.419)	(0.310)	(0.699)	(0.526)
$r_2 < r_2^c$ [+]	3.184***	2.971***	3.438***	3.240***	2.903***	2.582***
	(0.307)	(0.231)	(0.608)	(0.474)	(0.925)	(0.691)
Child Fixed Effects	no	yes	no	yes	no	yes
Observations	491		264		231	
(Low, upper) bounds	(4,8)		(2,6)		(0,4)	

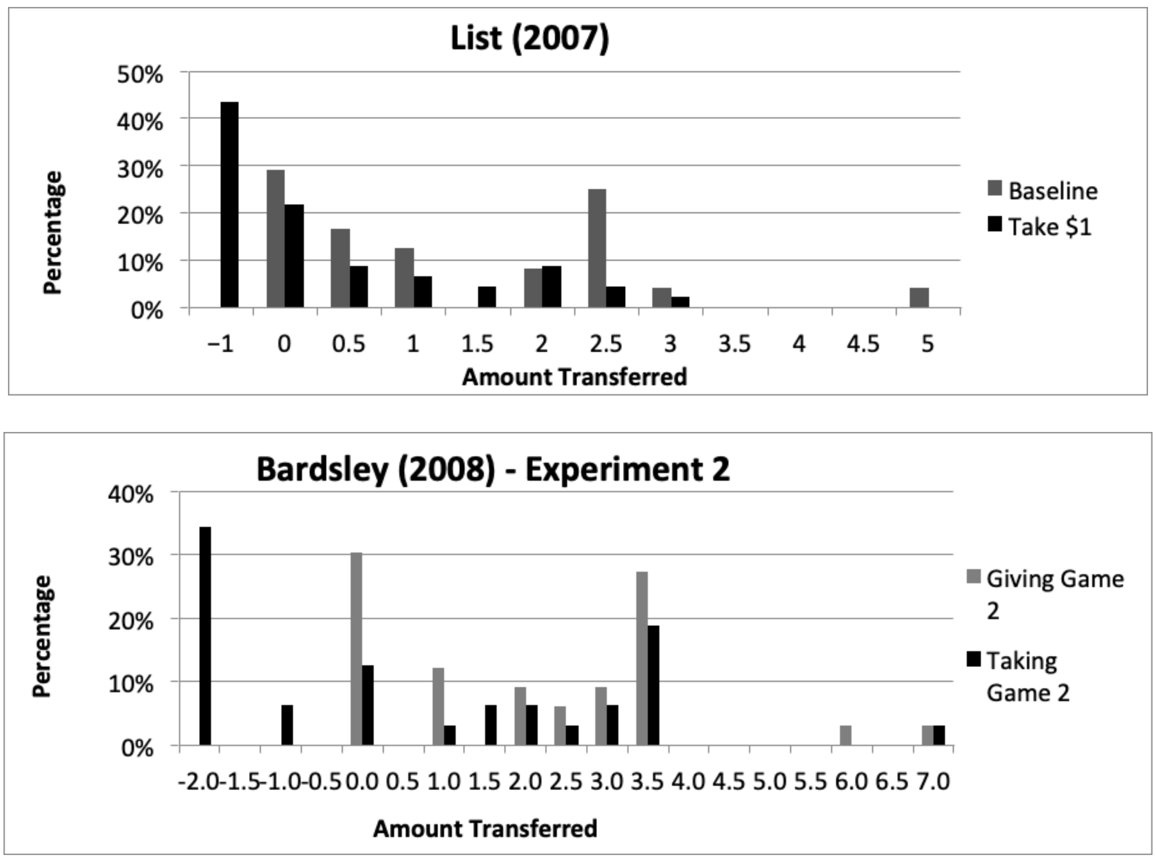
Notes: MMA predicted sign in square brackets. Entries are Tobit estimated coefficients. Standard errors in parentheses. *** p<0.01, **p<0.05, *p<0.1.

Table 3. **Effects of Recipient Moral Reference Dimension (Undergraduate Experiment; between subjects)**

Dep. Var: Dictator Payoff	r_1=15		r_1=19		r_1=11	
$r_2 < r_2^c$ [+]	-1.907	-1.653				
	(1.434)	(1.451)				
$r_2 > r_2^c$ [-]	-7.135***	-6.910***	-1.661*	-1.563*	-1.318**	-1.310**
	(1.667)	(1.681)	(0.858)	(0.885)	(0.618)	(0.605)
Demographics	no	yes	no	yes	no	yes
Observations	207	207	147	147	131	131

Notes: MMA predicted sign in square brackets. For r_1=15 columns: r_2^c=7 and r_2=3 in the first row and r_2=11 in the third row. For two r_1=19 columns: r_2^c=3 and r_2=7 for the third row. For two r_1=11 columns: r_2^c=7 and r_2=11 for the third row. Entries are average marginal effects. Demographics include gender, race, GPA, religion, major and study year. Standard errors in parentheses. *** p<0.01, **p<0.05, *p<0.1.

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

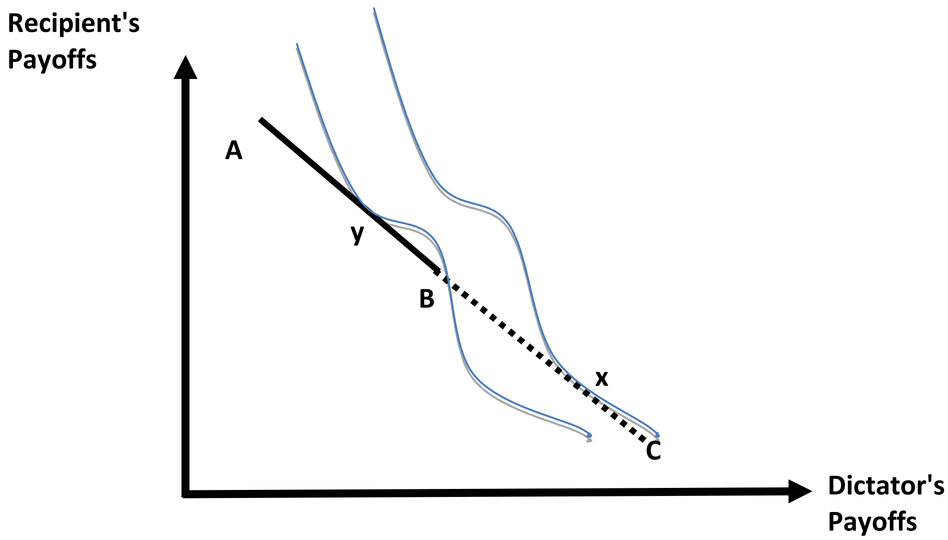


Figure 3. Endowments, Average Choices, and Moral Reference Points for Korenok et al.

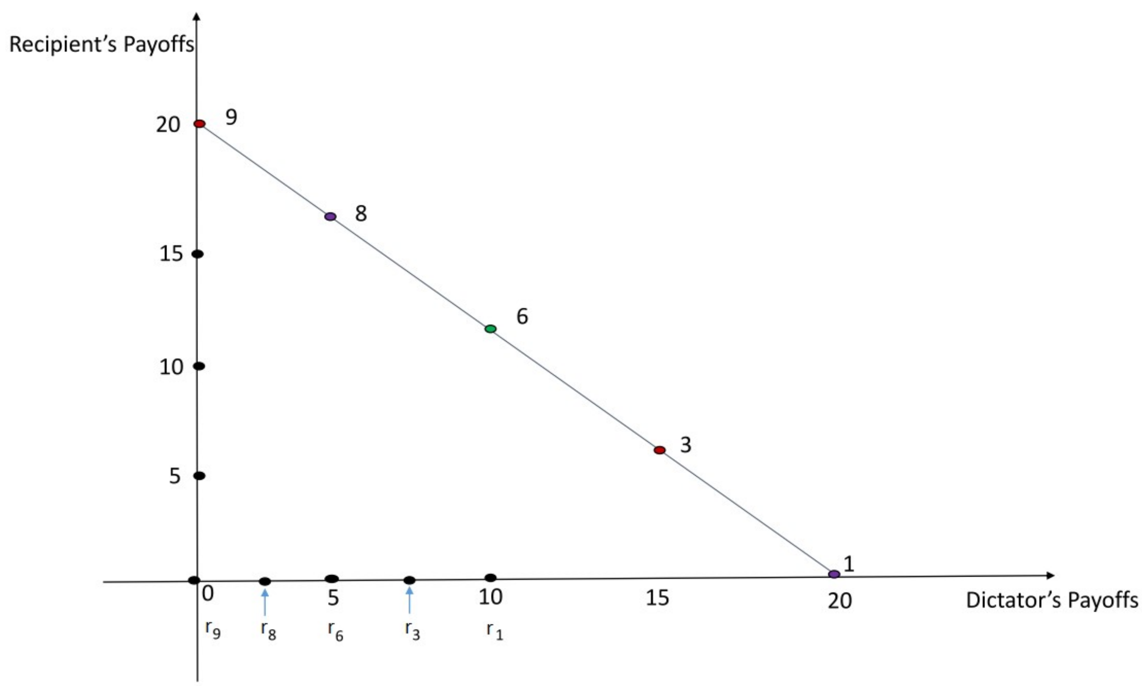


Figure 4. Example of Moral Reference Points in Dictator Games

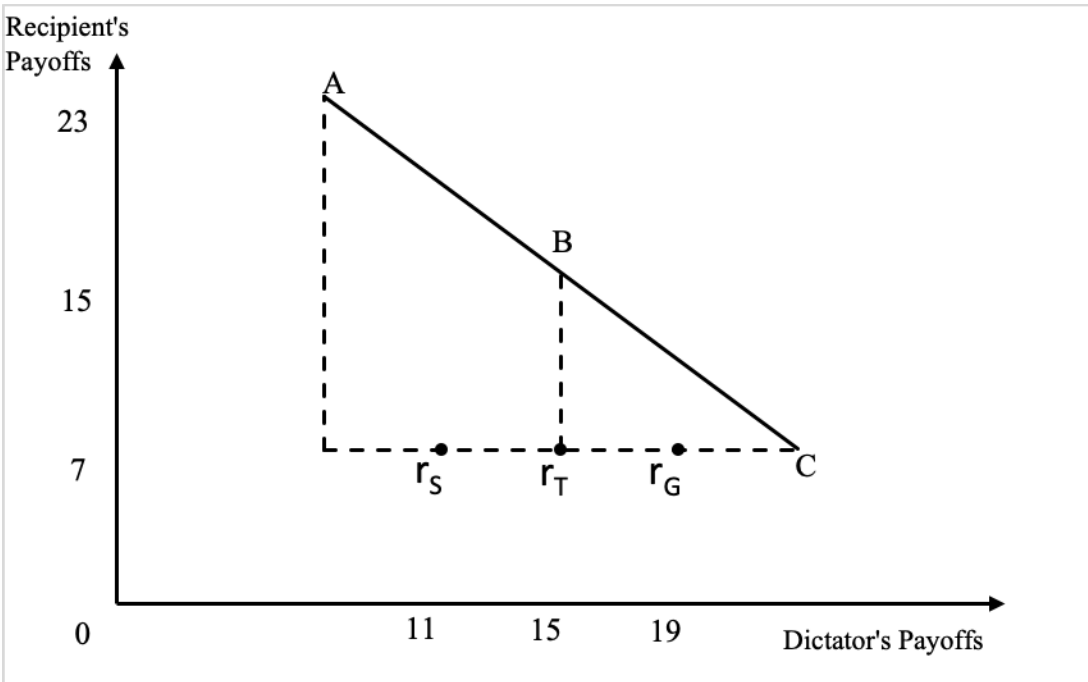


Figure 5. MMA and WARP for the Andreoni and Miller Experiment

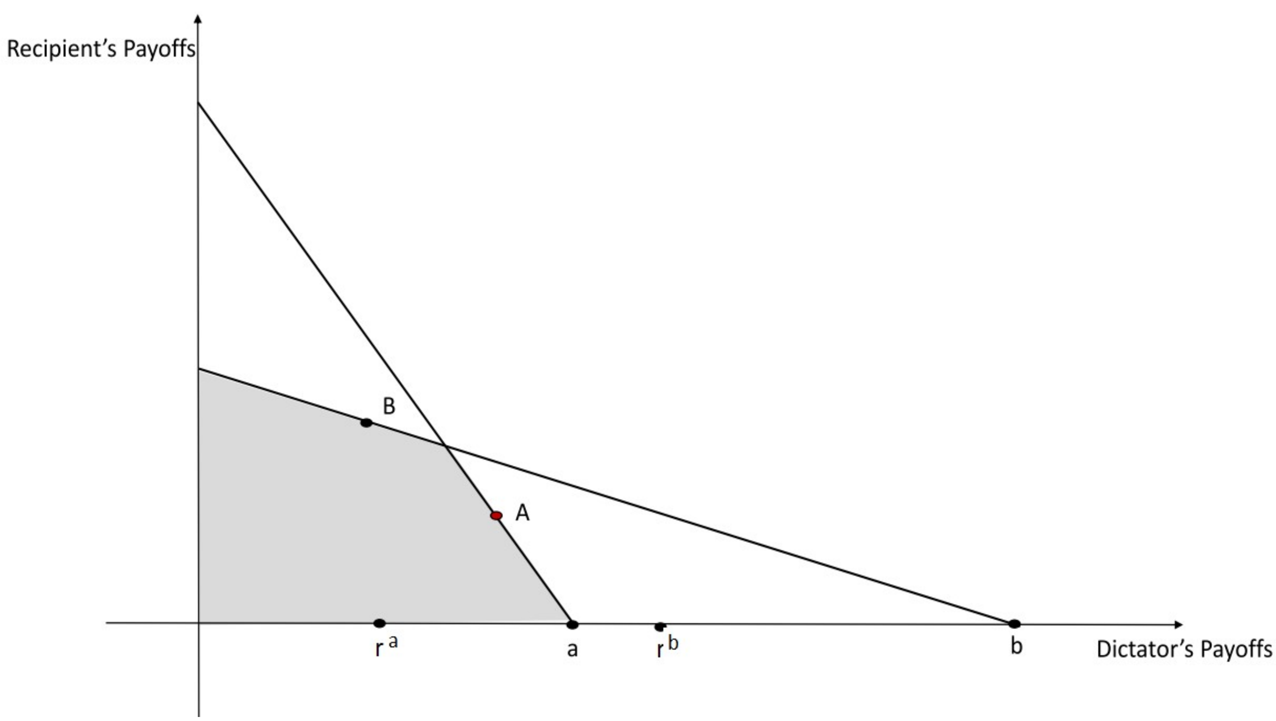
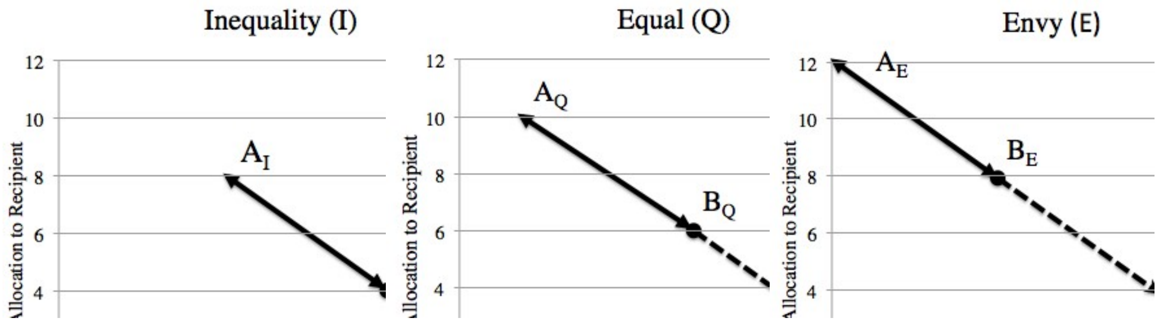


Figure 6. Children Exp. Feasible Sets: [A, B] for Give or Take, [A, C] for Symmetric

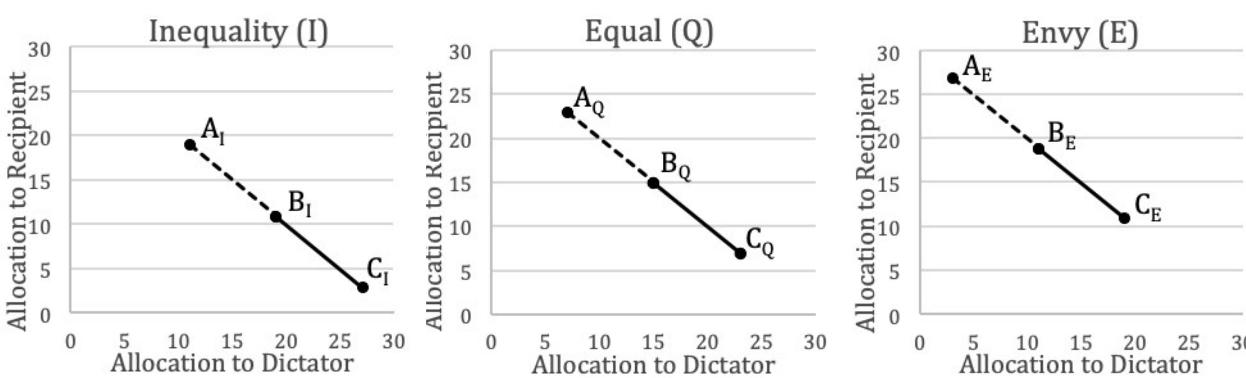


Note: This figure shows the feasible budget available for each treatment and action set. Participants in the Give or Take action sets can choose from [A, B], while participants in the Symmetric action set can choose from [A, C].

Figure 7: Experimental Setup for Children Experiment

Note: This figure displays the experimental environment. At left, blue and red plates indicate payoffs for the dictator and recipient. Stickers on top of the plate are variable endowment, while clearly visible stickers inside the boxes are fixed endowment. At right, a child participates in the experiment one on one with an experimenter.

Figure 8. Students Exp. 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.

ONLINE APPENDICES

Appendix A. Proof of Proposition 1

Let $\boldsymbol{f}$ belong to both F^* and G. Consider any $\boldsymbol{g}$ from G^*. As G and F have the same moral reference point, $\boldsymbol{r}^g = \boldsymbol{r}^f$, 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 $\boldsymbol{g} = \boldsymbol{f}$; i.e., f belongs to G^* which concludes the proof for Property α_M. Note, though, that any choice $\boldsymbol{g}$ in G^* must coincide with $\boldsymbol{f}$, 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 $\boldsymbol{y}$ 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

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

The minimal expectations point, $\boldsymbol{y}_*^F$ is defined as follows. For each player j, define player i's minimal expectation payoff with respect to j as

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

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 i^{th} element of the vector $\boldsymbol{y}_*^F$.

Appendix C. An example of MMA choices in dictator games with Outside Options (Lazear et al. (2012) experiment)

Here we provide an example using a simple 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.