7-1-2017

Moral Costs and Rational Choice: Theory and Experimental Evidence

James Cox
Georgia State University

John List
Georgia State University

Michael Price
Georgia State University

Vjollca Sadiraj
Georgia State University

Anya Samek
University of Southern California

Moral Costs and Rational Choice: Theory and Experimental Evidence

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

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

ABSTRACT

The literature exploring other regarding behavior sheds important light on interesting social phenomena, yet less attention has been given to how the received results speak to foundational assumptions within economics. Our study synthesizes the empirical evidence, showing that recent work challenges convex preference theory but is largely consistent with rational choice theory. Guided by this understanding, we design a new, more demanding test of a central tenet of rational choice—the Contraction Axiom—within a sharing framework. Making use of more than 300 dictators participating in a series of allocation games, we show that sharing choices violate the Contraction Axiom. We advance a new theory of moral reference points that augments standard models to explain our experimental data. Beyond capturing the data patterns in our experiment, our theory also organizes the broader sharing patterns in the received literature and has applications to strategic games with contractions.

JEL Classifications: C93, D03, D64
Keywords: experiment, giving, taking, altruism, moral cost

* 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. Constructive criticisms by referees and the editor motivated significant improvements in the paper.

1. INTRODUCTION

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

Within this line of research, a class of experiments is used to measure pro-social preferences, with typical 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, what is largely missing is a deeper understanding of whether individual choices violate deeply held economic tenets. At this point, it is too early to conclude definitively, but the received literature suggests that observed sharing behaviors are consonant with neoclassical theory. For instance, in a seminal study, Andreoni and Miller (2002) show that in a modified dictator game subjects' choices satisfy the key axiom of revealed preference theory.[1] More recently, Andersen et al. (2011) provide data that reveals demand curves for fairness in an ultimatum game are downward sloping. While in its infancy, this work suggests that certain sharing behaviors can be captured by the standard economic model.

The shortage of work testing 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, 1995; and 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

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

alternatives do a satisfactory job of organizing behavior. Choi et al. (2007) extend this analysis by developing an experimental protocol that allows the researcher to both test the consistency of choices with the assumption of utility maximization and estimate a two-parameter utility function for each individual. These examples are not exhaustive, as there are many other active research inquiries in this spirit, including those exploring intertemporal choice (see, e.g., Laibson, 1997; O'Donoghue and Rabin, 1999, 2001; Frederick et al., 2002), asymmetry and transitivity of preferences (Tversky, 1969; Slovic 1995; Cox and Grether, 1996; List, 2002), and conditional altruism (Dufwenberg and Kirchsteiger, 2004; Cox, Friedman, and Sadiraj, 2008).

Our study follows the spirit of this broader literature by exploring whether basic economic tenets are satisfied in sharing choices as observed in the dictator game, which has emerged as a workhorse in the social science literature. Recently, 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). 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.

The first goal of our study is to step back from the burgeoning literature and attempt to synthesize what we have learned theoretically from the experimental exercises of List (2007) and others. We explain that the traditional dictator game, wherein more than 60 percent of dictators pass a positive amount of money, is consistent with neoclassical convex preference theory (Hicks, 1946; Samuelson, 1947). Yet, more recent results from this literature (e.g., List, 2007; Bardsley, 2008; Cappelen, et al., 2013) provide evidence that challenges convex preference theory. Nevertheless, these new data are largely consistent with rational choice theory (Sen, 1971).

Our second goal is to build on the experimental literature by conducting a dictator game experiment that generates a stark test of a foundational assumption within economics: the Contraction Axiom.[2] For singleton choice sets, the Contraction Axiom (Chernoff, 1954) is the necessary and sufficient condition for a choice function to be rationalizable by a complete and

[2] For singleton choice sets, the Contraction Axiom states that if *x* is chosen from feasible set *F* then it will also be chosen from any contraction of set *F* that contains *x*.

transitive ordering (Sen, 1971).[3] To test whether this central theoretical condition holds, we present an experiment with dictator games in which we systematically vary both the feasible set and the actions available to the dictator.[4] Designing an experiment that preserves the feasible set but allows dictator giving or taking provides one type of test of the Contraction Axiom. This important design departure from the List (2007) and Bardsley (2008) work makes it possible for us to explore rational choice theory at a deeper level. Furthermore, by preserving the initial endowment but contracting the feasible set, we depart from recent literature on effects of social norms on play in dictator games (Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2015). Unlike this related work, our design allows discrimination between the effects on choices of initial endowments from the effects of contracting feasible sets while preserving endowments, which turns out to be crucial to discriminating between rational choice theory (the Contraction Axiom) and its special case, convex preference theory.

The experimental data yields several insights. First, we find that our subjects – students at Georgia State University - exhibit patterns of giving and taking behavior similar to other university students (List, 2007; Bardsley, 2008; Korenok, Millner, and Razzolini, 2014) and to a representative sample of Danish adults (Cappelen, et al., 2013). In particular, contracting the feasible set to remove taking options causes subjects to provide higher payoffs to recipients and keep less for themselves. Second, and most importantly, in our experiment such contraction causes subjects to keep more for themselves even when the contracted set contains the originally-chosen allocation, which is inconsistent with the Contraction Axiom and therefore at odds with extant rational choice theory. Crucially, combined with previous results, our data suggest how rational choice theory can be modified to explain the overall behavioral patterns.

This deeper understanding leads to the third goal of our study: to develop an axiomatic foundation for other-regarding behavior and test empirical validity of the new axiomatic theory with data from our own experiment and from previous work.[5] Our theoretical development follows the approach in Cox and Sadiraj (2010) to extend choice theory to accommodate dictator game data that violates a central tenet of conventional theory – in this case, the Contraction Axiom. The key component of our theory is the identification of moral reference points that are observable

---

[3] For set-valued choice functions, rationality is equivalent to Sen's (1971) Properties α and β (see below). Property α is the Contraction Axiom.
[4] In this paper, we have elected to use "action set" to refer to actions of taking or giving whereas "feasible set" denotes the conventional set of feasible allocations, i.e., it is ordered pairs of dictator's and recipient's payoffs.
[5] See also experiments by Grossman and Eckel, 2015, Engel, 2011; Korenok et al., 2013; Korenok et al., 2014; Zhang and Ortmann, 2014.

features of the environment, i.e., feasible sets and initial endowments.[6] We view our study as fitting in nicely with the "theory speaking to experiment and experiment speaking to theory" research culture that has permeated experimental economics for decades.

The remainder of our paper is structured as follows. Section 2 presents the design of our experiment and the procedures. Section 3 discusses the implications of extant theory and develops our axiomatic theory incorporating moral reference points. Section 4 presents our experimental results. Section 5 presents implications of our theory for related experiments in Andreoni and Miller (2002), Korenok, et al. (2014), Krupka and Weber (2013), and Lazear, Malmendier, and Weber (2012). Section 6 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 7 concludes.

2. BACKGROUND, DESIGN, AND PROTOCOL

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

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, my preferences may be defined over my hotdogs and my hamburgers. But the same formal theory of preferences can be applied to two commodities identified as my hamburgers and your hamburgers. Identification of the commodities in a bundle is an interpretation of the theory. In this way, neoclassical preference theory (Hicks, 1946; Samuelson, 1947) can be used for agents who are either self-regarding or other-regarding. As such, strictly speaking, the received results of generous sharing in standard dictator games do not represent a rejection of neoclassical preference theory. Rather, they represent a rejection of a joint hypothesis: neoclassical preferences and the assumption that preferences are self-regarding.

[6] 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.

More recently, List (2007) and Bardsley (2008), amongst others, have used laboratory dictator game experiments to explore how choices are influenced by introducing opportunities for the dictator to take from another subject. This line of work presents a challenge for convex preference theory, as we explain. We use this literature as our starting point, and design treatments that pose a more fundamental challenge to choice rationality than heretofore explored.

2.1 Experimental Design

Following List (2007), our design begins by introducing an action set in which the dictator can either give to or take from the recipient's initial endowment and compares outcomes in this augmented game to those observed in dictator games in which the participant can only give to, or take from, the recipient. We extend this line of inquiry by considering treatments that vary the initial endowments but preserve the permissible set of actions.[7] If the motivation behind choices is driven by final allocations only, as assumed in conventional theory, variation in the initial endowments within a given feasible set should have no impact on observed dictator behavior.

Figure 1 shows three budget lines labeled "Equal," "Inequality," and "Envy." The finite feasible sets are ordered pairs of integers on the lines. Labelling of the feasible sets reflects the location of the midpoints B_j, $j = Q, I, E$ 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.

The experimental design is 3 × 3: (Equal, Inequality, Envy) × (Symmetric, Take, Give).[8] In the Inequality-Give treatment (with endowment at point C_I in the middle panel): the recipient has an endowment of 3; the dictator has an endowment of 27 and can give up to 8 to the recipient.

---

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

[8] The treatments used in the experiment reported herein are similar to ones used in the experiment reported in the working paper, Cox et al. (2016), which utilizes variations in initial endowments and feasible actions to explore the importance of moral reference points on the choices of young children.

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

FIGURE 1 ABOUT HERE: FEASIBLE SETS

In the Inequality-Symmetric and Envy-Give treatments, the dictator is faced with an allocation decision over a budget set 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 feasible set for the Equal-Take treatment lies on and below the 45 degree line whereas the feasible budget set for the Equal-Symmetric treatment crosses the 45 degree line.

The nine treatments are constructed to stress-test all consequentialist theories: (a) the same action (of give or take) amount x produces very different allocations (consequences) in different treatment cells; (b) the same allocation (consequence) results from different give or take actions in different treatment cells.

2.2 Protocol

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

The two decision tasks each subject faced allow us to conduct within-subjects tests of consistency with rational choice theory. 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 so there would be no treatment order effects. 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.

3. THEORY DEVELOPMENT

As noted above, Figure 1 portrays the feasible sets faced by subjects in our dictator games. Self-regarding (or *homo economicus*) preferences imply the choice of C_j in all treatments. Models of other-regarding preferences predict choices that may differ from C_j. For strictly convex preferences, one can make additional statements as follows. When (in a Symmetric action set) a most preferred allocation $\mathrm{Q}_{\mathrm{j}}^{*}$ in feasible set $[A_j, \mathrm{C}_j]$ does *not* belong to the subset $[B_j, C_j]$ then strict convexity requires B_j to be the unique most preferred allocation when (in a Give or Take action set) the budget set is $[B_j, C_j]$.[9] In addition, when (in a Give or Take action set) a most preferred allocation $\mathrm{Q}_{\mathrm{j}}^{*}$ in feasible set $[B_j, \mathrm{C}_j]$ is not B_j then strict convexity requires Q_j^* to be the unique most preferred allocation when the budget set is $[A_j, C_j]$. These statements apply to neoclassical theory of other-regarding preferences and to popular models of social preferences

[9] For any given feasible allocation, X from $[B_j, C_j]$, allocation B_j is a convex combination of X and $\mathrm{Q}_{\mathrm{j}}^{*}$ (that belongs to $(B_j, C_j]$. Since $\mathrm{Q}_{\mathrm{j}}^{*}$ is revealed preferred to X in $[A_j, \mathrm{C}_j]$, by strict convexity B_j is strictly preferred to X.

(e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002).[10]

Choice implied by convex preferences and rational choice are different concepts. Convex preference theory is a special case of rational choice theory that imposes far stronger restrictions on observable choices. Consequentialist rational choice theory requires that there be a choice function (Zermelo, 1904) which satisfies certain consistency axioms (Samuelson, 1938; Chernoff, 1954; Arrow, 1959; Sen, 1971, 1986). A choice function, defined on a collection X of nonempty sets, assigns to each set S in X a choice set S^* of elements of S. A choice function is single-valued if the choice sets are singletons. If we let F^* denote the choice set when the opportunity set is F and G^* be the choice set when the opportunity set is G then the Chernoff (1954) Contraction Axiom (also known as Property α from Sen 1971) states:

Property α**:** if $G \subseteq F$ then $F^* \cap G \subseteq G^*$

In other words, a most-preferred allocation $f^* \in F^*$ from feasible set F remains a most-preferred allocation in any contraction of the set that contains the allocation f^*. For non-singleton choice sets, there is a second consistency axiom, Sen's (1971) Property β.[11] In this paper we consider singleton choice sets, in which Property α simplifies to: if $f^* \in G$ then $f^* = g^*$, i.e., f^* is also the choice in G.

For finite feasible sets, Property α (the Contraction Axiom) is the necessary and sufficient condition for singleton choice sets to be rationalizable by a weak order (Sen, 1971).[12] So we focus on properties of singleton choice sets and implications of Property α, which in our dictator game experiment are as follows:

Choice Function Hypothesis. Given feasible set $[B_j, C_j]$, allocation P_j^* is chosen when the action set is Take and the endowment is at B_j if and only if it is chosen when the action set is Give and the endowment is at C_j.

Contraction Hypothesis. If allocation Q_j^*, chosen in the Symmetric action set with

[10] Note that these statements would not apply to some models of warm glow preferences, such as Korenok et al. (2013).

[11] Property β: if $G \subseteq F$ and $G^* \cap F^* \neq \emptyset$ then $G^* \subseteq F^*$. In other words, if the most-preferred set F^* for feasible set F contains at least one most-preferred point from the contraction set then it contains all of the most-preferred points of the contraction set. For finite sets, Properties α and β are necessary and sufficient conditions for a choice function to be rationalizable by a weak order (Sen, 1971). In case of choice sets being singletons, Property β is automatically satisfied.

[12] A weak order is complete and transitive.

feasible set $[A_j, \mathrm{C}_j]$, belongs to the subset $[B_j, C_j]$ then Q_j^* is also chosen when the action set is Give or Take and the feasible set is $[B_j, C_j]$.

The Choice Function Hypothesis is an immediate implication of choice sets being singletons: the choice set of feasible set $[B_j, \mathrm{C}_j]$ contains one element, P_j^*. This hypothesis is also implied by the Property α because any set (e.g. $[B_j, \mathrm{C}_j]$) is a subset of itself.

A behavioral interpretation of the Contraction Hypothesis as follows. Choice of Q_j^* belonging to $[B_j, C_j]$ when the feasible set is $[A_j, C_j]$ reveals that allocations in $[A_j, B_j)$ are less desirable alternatives than Q_j^*; therefore excluding them from the feasible set should not affect choice. For the sets in Figure 1, if a subject chooses an allocation in $[B_j, C_j]$ in the Symmetric game then she should choose the same allocation in the corresponding Take and Give games. Unlike convexity, Property α has no implication for choice from feasible set $[B_j, C_j]$ if the choice from $[A_j, C_j]$ is contained in $[A_j, B_j)$ – a distinction that has been overlooked in the prior literature.

Data from many dictator game experiments with giving and taking refute strict convexity, but choice rationality remains unclear. For example, consider Figure 2, which shows data from List (2007) and Bardsley (2008).[13] 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 (statistically) the same – a pattern that is clearly refuted by the data.[14]

FIGURE 2 ABOUT HERE: HISTOGRAMS FOR LIST AND BARDSLEY DATA

Convexity, however, is not a necessary condition for choice rationality, so comparison of these 29% and 65% figures does not allow the researcher to draw conclusions about choice rationality. An illustration of rational choices for non-convex preferences is shown in Figure 3. A

[13] The data for List (2007) are from the *JPE* online appendix.
[14] The initial endowments are the same in these two treatments hence we can discuss implications of convex preference theory for either payoffs or transfers.

dictator with such preferences would choose y from the set $[A,C]$ but switch to x (rather than B) when she faces set $[B,C]$.

FIGURE 3 ABOUT HERE: EXAMPLE OF CHOICE WITH NON-CONVEX PREFERENCES

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

The above properties of rational choice theory imply that each of the bars portraying fractions of choices of \$0 to \$5 in the Take 1 treatment should not be higher than the corresponding bar for choices in the Baseline. With the exception of the bar at \$1.50 (corresponding to two observations in the Take 1 treatment), the List (2007) data are consistent with rational choice theory. Similarly, data shown in Figure 2 from Experiment 2 conducted by Bardsley (2008) are inconsistent with convex preferences but are mostly consistent with rational choice theory; the bar at \$1.50 (2 observations) is the only inconsistency with rational choice theory in Experiment 2 data.

As we shall explain in Section 4, data from some of the treatments in our experiment are inconsistent with rational choice theory, which prompts interest in modification of the theory. We next turn our attention to that topic to provide the theoretical foundation for our experimental design and to aid in the interpretation of the data patterns discussed in the empirical results section that will follow.

3.2 Theory of Moral Reference Points

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 incorporates moral reference points that are *observable features* of feasible sets.

We begin with an intuitive discussion of moral reference points, and moral monotonicity in choice behavior, for the specific context of dictator games. We then provide a formal definition of a moral reference point to make clear that it is an *observable* feature of feasible sets. We subsequently formalize the definition of moral monotonicity in the form of an axiom that modifies (consequentialist) rational choice theory.

To get a feeling of the concept of moral cost, consider the feasible sets in the right (Envy) and the left (Inequality) budget sets in Figure 1. The lowest feasible payoff for the recipient is $11 in the Envy-Give treatment and $3 in the Inequality-Give treatment. Allocating the recipient a payoff of $11 may inflict moral cost in the Envy-Give treatment (where it results from the most selfish[15] possible action of giving $0) whereas that same allocation of $11 may deliver moral benefit in the Inequality-Give treatment (where it results from the most generous possible action of giving $8). It seems plausible to assume that moral cost decreases with the difference between payoff allocated to the recipient and the recipient payoff that would result from the most selfish feasible option of the dictator. We build this intuition into our model by assuming that one dimension of the moral reference point for sharing behavior is determined by the recipient payoff at the most selfish feasible action available to the dictator. This is dimension r_2 in our formal definition of moral reference points below.

Another intuitive feature of moral cost relates to the dictator's position. Consider, for example, the Inequality-Take and Envy-Take treatments. Allocating oneself a payoff of $19 in the Inequality-Take treatment results from the most generous feasible action of taking $0 from the recipient. In contrast, allocating oneself a $19 payoff in the Envy-Take treatment results from the most selfish possible action of taking the maximum of $8 from the recipient. It seems plausible to assume that moral cost increases with the difference between the amount of payoff allocated to oneself and the own-payoff that would result from the most generous action available. We build this intuition into our model by assuming that the other dimension of the moral reference point for

---

[15] In discussing dictator games in this section, we label choices in the way it is most commonly done in everyday conversation. A "most selfish" choice is the one that provides the dictator with the largest money payoff (and the recipient with the smallest money payoff) out of all feasible choices. A "most generous" choice is the one that provides the recipient with the largest money payoff (and the dictator with the smallest money payoff) out of all feasible choices.

sharing behavior is determined, in part, by the own-payoff that results from the most generous action available to the dictator. But the dictator's moral cost may also vary inversely with the entitlement provided by her endowment in the game. In the Inequality-Symmetric treatment, for example, allocating oneself a payoff of $19 results from the action of standing pat, and neither taking of giving anything. In contrast, in the Envy-Symmetric treatment allocating oneself a payoff of $19 results from the action of taking the maximum feasible amount of $8 from the recipient. We build these intuitions into the model by assuming that the other dimension of the moral reference point is a convex combination of the own-payoff that results from the most generous possible choice and the entitlement payoff provided by the dictator's endowment. The central results that follow are invariant to the mixing proportion used in the convex combination. For simplicity, we use the mixing proportion one-half. This is dimension r_1 in our formal definition of moral reference point below.

We now formalize these intuitions with an axiomatic model that follows the approach used in Cox and Sadiraj (2010). The idea is to require that choices from feasible sets that preserve moral reference points (defined below) satisfy Property α and to present a concept of moral reference points that are determined by observable features of feasible sets. Throughout our discussion in this section we use dictator games as an example to illustrate concepts but the model has more general applicability, as explained in section 6.

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 compact finite set F. Let m^o and y^o be the maximum feasible payoffs for the dictator and the recipient, that is

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

It is natural to think of the minimal expectations point, (m_*, y_*) as the dictator's payoff when the recipient's gets y^o and recipient's payoff when the dictator gets m^o, i.e.,

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

Moral cost may depend on the minimal expectations point and payoff entitlement from the decision maker's endowment. Therefore we propose as a moral reference point an ordered pair that agrees with the minimal expectations point on the second (recipient's) payoff dimension and is a convex combination of the minimal expectations point and the initial endowment e_m on the first (dictator's) payoff dimension. For dictator game feasible sets we consider, the moral reference points are given by:[16]

$$\boldsymbol{r}^F = ((\tfrac{1}{2} m_*(F) + \tfrac{1}{2} e_m), y_*(F))$$

An algorithm for locating moral reference points is provided here for the Give, Take, and Symmetric action sets and Equal, Inequality, and Envy endowment treatments shown in Figure 1. 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.

The moral reference points, $\boldsymbol{r}^F = (r_1^F, r_2^F)$ for our various treatments are shown in Table 1. For the Inequality treatment, for example, the moral reference points are (23,3) in Give, (19,3) in Take, and (15,3) in Symmetric.

TABLE 1 ABOUT HERE

We use action sets shown in Figure 1 to illustrate a behavioral interpretation of moral reference points. If we look at the Inequality treatment the recipient coordinate, r_2 (=3) of the moral reference points are the same across the three scenarios: Give, Take and Symmetric. In contrast the dictator coordinate, r_1 varies from 23 to 19 to 15, which reflects the changes in the dictator minimal expectation payoff or initial endowment. One would expect a dictator to feel more entitled to a larger own payoff as her moral reference coordinate decreases while the recipient's coordinate remains constant. On the contrary, if we look at the Inequality-Symmetric, Equal-Take and Envy-Give treatments, the dictator coordinate, r_1 (=15) of the moral reference point remains constant. In contrast, the recipient coordinate r_2 varies from 3 to 7 to 11, which reflects the changes in the recipient's minimal expectation payoff. One would expect a dictator to feel obliged to allocate more to the recipient as the recipient's moral reference coordinate increases while the dictator's coordinate remains constant, as formalized in MMA.

---

[16] A less specific definition of the moral reference point is $(r_1, r_2) = ((\theta m_*(F) + (1-\theta)e_m), y_*(F))$ where θ is between 0 and 1. Any value of $\theta \in [0,1)$ provides moral reference points that make MMA consistent with all of the contraction and action set effects we discuss for the experiment reported herein.

Our many applications of theory in this paper will all be to two-agent (dictator and strategic) games. Some future applications of our theory to novel situations, however, will require identification of the observable moral reference points when there are more than two agents. 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_i^0 $(i = 1, 2, \cdots, n)$ be the maximum feasible payoff for player i $(= 1, 2, \cdots n)$, 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 $S_i = \{y_{*ij}^F : j ¹ i)$ 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 $\boldsymbol{y}_*^F$.

We now turn our attention to moral monotonicity. We postulate that agents' choices, characterized by moral cost concerns, satisfy a monotonicity criterion with respect to moral reference points. We now formalize this in an axiom for n-players. Let F^* be the choice set for feasible set F (and similarly for G^* and G). Let $\boldsymbol{r}^G$ and $\boldsymbol{r}^F$ be the moral reference points for feasible sets G and F, and let $\langle\rangle$ be the notation for "not smaller" or "not larger." For every player i $(= 1, \cdots, n)$ one has:

Moral Monotonicity Axiom (MMA):

If $G \subseteq F$, $r_i^G \langle\rangle r_i^F$ and $r_{-i}^G = r_{-i}^F$, then $\boldsymbol{f}^* \in F^* \cap G \Rightarrow g_i^* \langle\rangle f_i^*, \forall \boldsymbol{g}^* \in G^*$

In words, MMA says the following. Suppose that G is a subset of F that contains some choice $\boldsymbol{f}^*$ from F. Suppose also that the moral reference points of F and G differ from each other only with respect to the value of dimension i. If the moral reference point in G is more favorable to individual i, then no choice from G allocates him less than $\boldsymbol{f}^*$. Similarly, if the moral reference point in G is less favorable to player i, then no choice from G allocates him more than $\boldsymbol{f}^*$.

What are the implications of MMA for contractions that preserve moral reference points and contain choices from the bigger set? We show that for such subsets MMA implies that the choice set is a singleton and that conventional axioms of rationality (Sen's 1971 Properties α and β) are satisfied. The modified form of Sen's Property α for sets that preserve the moral reference point is[17]

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 simplifies to: if $\boldsymbol{f}^* \in G$ and $\boldsymbol{r}^G = \boldsymbol{r}^F$ then $\boldsymbol{g}^* = \boldsymbol{f}^*$ is chosen in G. We are ready now to state implications of MMA for choices.[18]

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.

3.3 Testable Implications of MMA vs. Property α

MMA has many testable implications for the action sets and endowment treatments in our experiment. For example, in the Equal treatment, the moral reference point is (19,7) in the Give action set but (15,7) in the Take action set. Therefore, MMA implies that the choice in Give is southeast of the choice in Take, which means the dictator allocates a (weakly) larger own payoff in Give than in Take. This contrasts with the implication of conventional rational choice theory that the Give and Take action sets have the same outcomes. The same type of argument can be used to show that, in the Envy and Equal endowment treatments, MMA implies a smaller allocation to the dictator in the Take than in the Give (but larger than in the Symmetric) whereas conventional theory implies identical allocations when from [B,C].

[17] 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 $G^* \subseteq F^*$.

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

4. EXPERIMENTAL RESULTS

4.1 Overview

612 subjects (306 dictators) participated in the experiment. None of the dictators had previous experience in dictator games. Each session lasted approximately 50 minutes and each dictator made two decisions. The actual payoffs (from the randomly selected payoff rounds) for dictators were: $19.46 (average) with the range $8 (minimum) to $27 (maximum). Average payoffs and transfers[19] for all data from nine treatment cells are reported in Table 1.

Less than 1/3 (166 out of 612) of observed choices correspond to the most selfish feasible options and less than 1/5 (57 out of 306) of dictators appear selfish (i.e., always choose the most selfish option). Data exhibit egocentric altruism (Cox and Sadiraj, 2007) as almost all choices (98%) are such that the dictator's final payoff (weakly) exceeded recipient's final payoff.[20] All data from Give and Take treatments with feasible sets $[B_k,C_k]$ are usable for testing convex preference theory, conventional rational choice theory (Property α), and modified rational choice theory incorporating the Moral Monotonicity Axiom (MMA). Choices from $[A_k,B_k)$ in a Symmetric treatment are not usable for testing rational choice theory. Overall, 94% (575 out of 612) of choices are from sets $[B_i,C_i]$ and are thus usable in testing all theories.

We begin with tests of convex preferences, Principal α and MMA using only within-subjects choice pairs. Subsequently, we use all of the data to ascertain whether observed subjects' transfers are affected by moral reference points as predicted by MMA.

4.2 Consistency of Dictators' Choices with Theoretical Models.

Each dictator made two decisions from the same (Inequality, Equal or Envy) environment: 96 dictators faced budget set $[B_k,C_k]$ twice, in one Give action set and one Take action set; a different group of 98 dictators faced budget set $[A_k,C_k]$ in the Symmetric action set and budget set $[B_k,C_k]$ in the Give action set; and another group of 112 dictators faced budget set $[A_k,C_k]$ in the Symmetric action set and budget set $[B_k,C_k]$ in the Take action set. We created a dummy variable, Consistency, that takes value 1 only if a individual's two decisions agreed with a theoretical prediction. For each dictator, we construct three distinct consistency measures capturing whether choices for that

[19] "Transfer" is defined as the amount by which the recipient's payoff exceeds her minimum expectations payoff. In a Give treatment, the transfer is the amount the dictator gives to the recipient. In a Take treatment, the transfer is the amount the dictator does *not* taken from the recipient.

[20] Dictator's final payoff was strictly larger than the recipient's payoff in 80% (489 out of 612) of choices.

subject were consistent with the predictions under (i) MMA, (ii) Property α, and (iii) convex preference theory.

To test whether choices in our data are better explained by MMA than either of our alternative models, we conduct a sign test using our indicator variable Consistency. Specifically, for each agent *i* we construct a new variable z_i that is the difference in the Consistency indicator under the assumption that choice reflects MMA and the indicator for each of the alternative models. We then drop individuals whose choices are consistent with both MMA and the alternative model and assign a value of -1 to dictators whose choices are inconsistent with both models. Under the null hypothesis that MMA is no more likely to organize choice in our experiment than either of the alternative models, this new variable should follow a binomial distribution and the likelihood that $z_i = 1$ should be one-half. The alternative hypothesis is a one-sided test that the likelihood $z_i = 1$ is greater than one-half.

Table 2 reports the fraction of dictators whose choices are consistent with a given model – convex preference theory, MMA, or Property α. The first three rows of the table correspond to the choices for dictators who were assigned to a given environment (i) Inequality, (ii) Equal, or (iii) Envy whereas the final row reports data for all dictators in our experiment. The first two columns of the table compare whether observed choices are better organized by MMA or convex preference theory whereas the final two columns compare whether choices are better organized by MMA or standard rational choice theory.

Property α vs. MMA. We test for data consistency with Property α **(**the Contraction Axiom) by excluding 37 (out of 306) dictators whose choices in a Symmetric action set were from $[A_i,B_i)$ because Property α makes no prediction for their choices when the budget set is $[B_i,C_i]$. The last two columns of Table 2 show the fraction of choices for the remaining 269 dictators that were consistent with Property α (column 3) and MMA (column 4). As noted in the final row of Table 2, about half of the remaining dictators (135 out of 269) made choices that are consistent with Property α and 78% (210 out of 269) made choices that are consistent with MMA. The Sign test weakly rejects Property α in favor of MMA using data from all three environments. The observed pattern, whereby the fraction of dictators whose choices are consistent with MMA is greater than the fraction whose choices are consistent with standard rational choice theory, is robust across the three (Envy, Inequality and Equal) environments. However, only data from the Equal environment reject Property α in favor of MMA at conventional levels.

Convex Preferences vs. MMA. Convexity requires invariance to action sets for all choices from $(B_i, C_i]$ but it makes no predictions on $[A_k,B_k)$ when B_k is chosen from $[B_k,C_k]$. Eighteen dictators in our experiment chose B_k in the Give or Take scenario and an allocation in $[A_k,B_k)$ in the Symmetric action set. Since these choice pairs are uninformative for testing convexity they are excluded in the analysis below, leaving us with 288 dictators.[21] Convexity predicts the dictators' two payoffs in their two choices will be the same whereas MMA predicts a larger payoff in the scenario with the larger dictator moral reference point dimension r_1. The first two columns of Table 2 show the fraction of choices for the 288 dictators whose choices are consistent with convex preference theory (column 1) and MMA (column 4). As noted in the final row of the table, less than half (47% or 135 out of 288) of dictators make choices that are consistent with convex preferences; 80% (229 out of 288) are consistent with MMA whereas 20% violate MMA. The first two columns of Table 2 show consistency figures for Convexity and MMA for each pair of games as well as pooled data (last row). Convexity is rejected in favor of MMA by the Sign test at conventional levels of significance for the pooled data and for dictators assigned to either the Inequality or Equal environments.

TABLE 2 ABOUT HERE. CONSISTENCY WITH CONVEXITY, PROPERTY α AND MMA

Size Effects and MMA. Recall that the recipient moral reference point dimension r_2 is fixed for any given (Inequality, Equal or Envy) environment (see Table 1) so variation in choices that violate Property α or convex preferences reveal an r_1 effect. To further investigate this effect, a new variable $\triangle P$ was constructed by subtracting a dictator's observed payoff in the treatment with smaller MMA-predicted payoff from the dictator's payoff in the treatment with (strictly or weakly) larger MMA-predicted payoff. Figure 4 shows histograms of $\triangle P$ across the three treatments as well as pooled data. The null hypothesis, $\triangle P = 0$ is implied by Property α or convex preferences. The alternative hypothesis, $\triangle P > 0$ is consistent with MMA.

Figure 4 About Here. Histograms of ΔP (within subjects)

[21] An alternative way is to replace all choices from $[A_k,B_k)$ with B_k before analyzing data. Findings from this alternative procedure are similar (though more in favor of MMA) to excluding data from these (18) dictators, which is a more conservative approach.

The overall mean of $\triangle P$ (288 subjects) is 0.60 (95% C.I. is [0.260, 0.934]). Means of $\triangle P$ across the three games are: 0.23 (Envy, one-sided p-value=0.138), 0.96 (Inequality, one-sided p-value=0.005) and 0.63 (Equal, one-sided p-value=0.023).[22] The null hypothesis, $\triangle P = 0$ is rejected in favor of the alternative hypothesis implied by MMA using the t-test (one-sided p-value = 0.000).

To test for Property a we exclude choices of thirty-seven subjects for whom Property a makes no predictions. The mean of $\triangle P$ (269 subjects) is 0.19 (95% C.I. is [-0.096, 0.475]). Means of $\triangle P$ across the three games are: 0.05 (Envy, one-sided p-value=0.379), 0.12 (Inequality, one-sided p-value=0.343) and 0.40 (Equal, one-sided p-value=0.082). The null hypothesis, $\triangle P = 0$ is arguably rejected in favor of the alternative hypothesis implied by MMA using the t-test (one-sided p-value = 0.096).

Together, results from the above tests using within-subjects data, can be summarized as follows.

Result 1: Data support MMA when the moral reference point is more favorable to the dictator. Specifically, our data suggest that final allocations depend upon the dictator's moral reference point and that payoffs for the dictator are increasing in this reference point. This finding is at odds with convex preference theory and standard rational choice theory but consistent with the predictions of MMA and our theory of moral costs.[23]

4.3 Test of MMA when the Recipient Moral Reference Dimension Changes

MMA says that the recipient's payoff increases in r_2 when r_1 is fixed whereas convexity and Property α predict no effect from changes in r_2. To test for r_2 effects, we need to look across environments.[24] We have data for three levels of r_1 that can be used to test responses to changing

[22] For 18 subjects with choice B in a Give or Take action set and from $[A_i, B_i)$ in the Symmetric design, a positive $\triangle P$ *could* reflect the constraints of the experimental design rather than an r_1 effect. To prevent this potential confound from possibly biasing the test we are excluding $\triangle P$ values for these 18 subjects.

[23] Result 1 is consistent with findings from Cox et al. (2016) who use a similar experimental design to test the importance of moral reference points on the choices of young children. As in our experiment, data from Cox et al. (2016) show that final allocations depend on both initial endowments and feasible actions. As such, dictator choices in their experiment violate the standard model of rational choice and any model that assumes convex preferences but provide support for MMA.

[24] 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 r_2.

values of r_2. We use Tobit models for our formal analysis throughout since the feasible choices are bounded by the design of the experiment.

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. Under standard models, the estimated coefficient on the minimal expectations point should be equal to zero whereas MMA predicts that recipient payoffs are increasing in r_2 and thus a positive coefficient on this measure. Table 3 presents results for these models. The first two columns restrict the analysis to the subset of choice where r_1 = 15. The third and fourth columns restrict the analysis to those choices where r_1 = 19 and the final two columns to those choices where r_1 = 11.

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

Data for r_1 ***=19:*** Treatments Inequality-Take and Equal-Give have both r_1=19 but r_2 values are 3 and 7, respectively. The feasible payoffs for the recipient are from: [3,11] in

Inequality-Take and [7,15] in Equal-Give. Set [7,11] is a subset of both sets, therefore we run Tobit regressions of recipient's final payoff with low bound 7 and upper bound 11. Tobit estimates for effect of r_2 on recipient's payoff are positive ($p<0.1$) which weakly rejects the null hypothesis from conventional preferences that the estimate is 0 in favor of the alternative hypothesis from MMA.

Data for r_1 ***=11:*** There are two treatments (Envy-Take, $r_2 = 11$ and Equal-Symmetric, $r_2 = 7$) with $r_1 = 11$. The feasible set, [11, 19] in Envy-Take is a contraction of the feasible set, [7,23] in Equal-Symmetric. So, we run Tobit regression with bounds 11 and 19. Tobit estimates for r_2 are positive ($p<0.05$), which rejects the null hypothesis from conventional preferences that the estimate is 0 in favor of the alternative hypothesis from MMA.

Thus, using between-subjects data we conclude that:

Result 2: Data are consistent with MMA when the moral reference point is more favorable to the recipient.

Across all models, the estimated coefficient on r_2 (the recipient's minimal expectations point) is positive. Holding the dictator's moral reference point constant, we find that a recipient's payoff is increasing in the minimal amount they could earn given the underlying budget set. This dependence is at odds with standard models, but is consistent with the predictions of MMA and provide additional evidence that moral reference points influence dictator behavior.

4.4 MMA and Transfers

The previous two sections provided direct tests of MMA with the focus on the payoff of the player favored by the moral reference point. In this section, we turn our attention to indirect implications of MMA and the effect of moral reference points on transfers, which is the dictator's choice defined in terms of "giving". [25] Unless we look at budget sets with certain characteristics (as in the previous sections on direct tests of MMA) differences in the support of feasible budget sets across environments confounds our ability to use payoffs to test additional implications of MMA.[26] To see the problem, note for example that the dictator's payoff that corresponds to the most selfish

---

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

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

feasible choice decreases from $27 (Inequality) to $23 (Equal) and down to $19 (Envy). By design, payoffs to dictators in the Envy treatments will be lower than those in the Inequality treatments independent of any choice they make. Looking at transfers (rather than payoffs) offers a way to control for this confound as the set of feasible transfers is invariant across our three environments (Inequality, Equality and Envy).

The feasible set of transfers is: [0,8] in both Give and Take action sets and [0,16] in the Symmetric action set for all three environments. Appendix B provides detailed derivations of the implications of MMA and conventional theory for the effects of changing r_1 and r_2 on transfers. However, the basic intuition underlying these formal derivations is as follows. In terms of r_2 and dictator's payoff *m*, any feasible transfer, t satisfies the equation (*) $t = 30 - r_2 - \mathrm{m}$.[27] Conventional theory (Property α or convex preferences) requires that the dictator's most preferred allocation, $\left(m^*, y^*\right)$ is preserved in all budget sets *that contain it*. Preservation of m^* requires that optimal transfer, t^* decrease in r_2 but be invariant with respect to r_1. In contrast, for MMA we have: (1) larger r_1 (ceteris paribus) implies larger m^* which comes with a smaller t^*; and (2) the direct effect (see (*)) of a larger r_2 on t^* is negative whereas the indirect effect, $-\triangle m^* / \triangle r_2 > 0$ is positive because a larger r_2 (ceteris paribus) increases y^*. The effect of r_2 on the optimal transfer is negative as the direct effect is stronger (see Appendix B).

Thus, we have the following testable hypotheses that allow us to evaluate whether transfers in our experiment are better organized by standard rational choice theory or our alternative model and MMA:

H_{r_1} : Marginal effects of r_1: 0 (Property α) or negative (MMA)

H_{r_2} : Marginal effects of r_2: -1 (Property α) or between -1 and 0 (MMA)

The mean transfers are 4.99 (Inequality, r_2 =3), 3.24 (Equal, r_2 =7) and 2.37 (Envy, r_2 =11).[28] This decreasing pattern is predicted by both MMA and conventional theory. However, the rate of decrease seems to be half of the size predicted by conventional theory. Table 4 reports results of a

---

[27] The recipient's final payoff is $\mathrm{y} = r_2 + \mathrm{t}$ in every treatment and as our games are zero-sum games, the dictator's payoff is $m = 30 - \mathrm{y} = 30 - r_2 - \mathrm{t}$.

[28] Kruskal-Wallis test: Chi-squared=30.25, p-value=0.001; for each subject, the data point is the mean of two observed transfers.

Tobit regression that allows us to estimate the effect of changing moral reference points on observed transfers. The list of regressors includes dictator's (r_1) and recipient's (r_2) coordinates of moral reference points of budget sets and, in model (2), demographic controls. As each dictator made two choices, we cluster standard errors. As a robustness check, Table 4 also present results from a Hurdle model (Cragg, 1971) which allows for the effects of moral reference points to differ along the extensive (whether to make a positive transfer) and intensive (the amount of any positive transfer) margins.

As noted in the first row of Table 4, the estimated coefficient on r_1 is negative and different from 0. The dependence of transfers on r_1 rejects conventional preferences in favor of MMA. The estimate of r_2, row two of Table 4, is negative which is consistent with the trend observed in above reported means of transfers across games. The Wald test rejects the conventional theory hypothesis that the estimate equals -1.[29] The estimates are consistent with MMA and are robust to both the inclusion of demographics in the regression and the use of a Hurdle model.

This provides our next result based on within-subjects and between-subjects data analysis.

TABLE 4 ABOUT HERE: TESTING r_1 and r_2 Effects on Transfers

Result 3: Convex preferences and Property α are rejected in favor of MMA.

In summary, our data provide empirical support for the predictions of changing moral reference points on transfers under MMA. In contrast, the data call into question the standard model of rational choice and models that assume convex preferences as organizing behavior in sharing games.

Alternative Models. We now briefly look at implications of our data for alternative models of behavior: random choices, selfish preferences, social preferences, reference dependence (Koszegi and Rabin 2006), and sharing and sorting (Lazear, et al. 2012).

Random Choice: Our dictator games are simple and, in addition our subjects participated in two practice rounds before making each of the two decisions. Nevertheless, if subjects are not paying attention any feasible transfer, any t is equally likely to be chosen. The hypothesis of random choice is rejected because parameter estimates for r_1 and r_2 are statistically significant.

[29] F(1,610)=30.50, p-value=0.000.

Selfish Preferences: Two-thirds of the transfers are positive and four out of five of our dictators made at least one positive transfer. Any feasible non-zero transfer reduces a dictator's payoff; therefore, this model predicts that the transfer is always 0, and hence, neither changes in r_1 nor r_2 will have an effect. Parameter estimates for both r_1 and r_2 are statistically significant; hence our data reject 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), and egocentric altruism (Cox and Sadiraj 2007) assume convex upper contour sets. Because our data reject convex preference theory, these social preferences models are also rejected.

Reference Dependent Model. Koszegi and Rabin (2006) develop a model of reference dependence that has recently seen a surge in applied work. Predictions of this model for our games are similar to standard rational choice theory because, in deterministic settings, optimal "consumption" derived for the conventional preferences model is the "preferred personal equilibrium" in the reference dependent model.[30] Because our data reject conventional theory, the reference dependent model is also rejected.

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

To summarize, not only do our data provide evidence at odds with standard rational choice theory, the data are also at odds with a suite of alternative behavioral models that have been used to explain sharing. Viewed in its totality, we thus believe our data provides compelling evidence that objectively defined moral reference points matter and influence choice in a manner consistent with MMA.

5. IMPLICATIONS OF MMA FOR OTHER TYPES OF DICTATOR GAMES

To formalize the ways in which moral reference points may influence decision-making in dictator games, we introduced the Moral Monotonicity Axiom (MMA) and applied it to analyze data from

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

our experiment. Yet, MMA has broader implications for choice in a range of related experiments including standard (give-only) dictator games (Andreoni and Miller 2002), other dictator games that compare the effect of give versus take actions on choices (Korenok et al. 2014), the "bully" dictator game (Krupka and Weber 2013), and dictator games with outside options (Lazear, Malmendier, and Weber 2012).

5.1 MMA and WARP

As previously mentioned, Andreoni and Miller (2002) conducted dictator game experiments that varied underlying budget sets and applied the generalized axiom of revealed preference (GARP) to analyze the consistency of choices in their setting. Figure 5 illustrates two budget sets like those that the dictator can face in the Andreoni and Miller design. Let point *a* denote the endowment on the steeper line and point *b* denote the endowment on the flatter line. Further, consider the shaded quadrilateral that is the intersection of sets bounded by the steeper and flatter budget lines. Viewed through the lens of MMA, the shaded quadrilateral set can be considered a feasible set with endowment at point *a*. The minimal expectations point is the origin (0,0) for all three feasible sets. Therefore, the moral reference points for the three feasible sets are on the horizontal axis, halfway between 0 and the respective endowment points. The moral reference point r^b for the budget set represented by the flatter budget line is more favorable to the dictator than the moral reference point r^a for the set represented by the steeper budget line.

FIGURE 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. Then MMA (see Proposition 1) requires that A also be chosen from the quadrilateral set because it is a contraction of the feasible set represented by the steeper line that preserves the moral reference point. Suppose that B is chosen from the lower flat triangle. MMA requires that the choice in the quadrilateral (which is also a contraction of the lower flat triangle) allocates to the dictator less than B does, because r^a is to the left of r^b. But this contradicts the choice of A from the quadrilateral set. Thus, any pair of choices of type A and B that violate WARP also violate MMA. In fact, MMA places tighter restrictions on data than does WARP (e.g., in Figure 5 WARP implies point A must be northwest of the intersection whereas MMA implies it must be west of point B).

5.2 Give and Take: MMA vs. Warm Glow

Korenok et al. (2014) report a dictator game experiment to test the theoretical model of warm glow developed by Korenok et al. (2013). In particular, the authors explore the effects of changing endowments and framing actions as giving to or taking from the recipient. Korenok et al. (2014) explain that data from their experiment is inconsistent with the predictions of their theory which, in this instance, are the same as the predictions of the conventional rational choice model.

Yet, the exhibited data patterns are consonant with our theory of moral costs.[31] Figure 6 illustrates five different scenarios in the Korenok et al. (2014) experiment. In all five scenarios, the feasible set is the same set of discrete points on the budget line shown in Figure 6. 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 6. 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 in 6.

FIGURE 6 ABOUT HERE: ENDOWMENTS & MORAL REFERENCE POINTS

The Korenok et al. (2013) theory and conventional rational choice theory both imply that choices will be invariant to changes in the endowments in the experiment. In contrast, our theory implies that choices will monotonically track changes in the underlying endowment points. To see this, note that the minimal expectations point is the origin (0,0) in all scenarios. Hence, the corresponding moral reference points for all scenarios are on the horizontal axis, halfway between 0 and the dictator's endowments for each of the respective scenarios. We have illustrated the various moral reference points in Figure 6 as r_j, for scenarios j = 1, 3, 6, 8, 9. MMA implies that choices monotonically move northwest as the endowment moves northwest along the budget line.

Let S1 ($4.05) denote the average payoff of $4.05 to the recipient in scenario 1. Using this same convention to reflect payoffs in the remaining scenarios, we have that the average recipient payoffs for the five scenarios are: S1 ($4.05), S3 ($5.01), S6 ($5.61), S8 ($6.59), and S9 ($6.31).

[31] Although we use the Korenok et al. (2014) data to explore implications of alternative theories, caution is called for in basing conclusions on those data because the payoff protocol used in the experiment is not incentive compatible. Their experiment involves role reversal in which each subject plays both dictator and recipient and is paid for both decisions. This payoff protocol might create an incentive for strategic behavior, not an incentive for truthful reporting of distributional preferences. Korenok et al. (2013), aware of this issue, report that this payoff protocol did not introduce significant bias in their experiment. Incentive compatibility of alternative payoff protocols is examined at length in Cox, Sadiraj, and Schmidt (2015).

The fact that average payoffs differ across endowments is inconsistent with predictions from the Korenok et al. (2013) theory and conventional rational choice theory. Importantly, however, the observed changes are as predicted by our theory except for the decrease from $6.59 to $6.31 between scenario 8 and scenario 9 – a difference that Korenok et al. report to be statistically insignificant at conventional levels.

5.3 MMA and Bully Games

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

Moreover, the distribution of elicited norms reported in Krupka and Weber's Table 1 are also consistent with MMA. A paired t-test of the two distributions rejects the null hypothesis of no effect (implied by Property α) 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 objectively defined moral reference points.

5.4 MMA and Outside Options

Lazear, et al. (2012) report an extended experimental design for dictator games that includes an outside option that allows subjects to opt out of the dictator game. Their Experiment 1 is a between-subjects design in which one group of subjects plays a "distribute $10" dictator game and another group of subjects can choose an outside option, that pays the dictator $10 and the other subject $0, or choose to play the distribute $10 dictator game.[32] 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

[32] 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)".

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.[33]

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 also consistent with our moral cost model in which choosing the outside option allows the decision maker to avoid moral costs from making the sharing decision whereas choosing to play the game involves this cost, as we now explain.

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

Here we provide an example using a simple utility function, $u(m, y) = m + \theta\sqrt{y}$. By sorting out, a subject can avoid the moral cost of making the sharing decision, obtain payoff allocation

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

(10,0), and utility $V(out) = 10 + q' \, 0$. 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

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

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

1. Any agent with (*) $\theta^2 < 4(15 - S)$ prefers sorting out and realizing payoff (10,0) to sorting in and being able to choose Pareto-dominating payoffs.
2. As S increases, inequality $S - 5 + \theta^2 / 4 > 10$ becomes more likely to be satisfied and therefore the fraction of subjects sorting in increases, as observed in Experiment 2.

Experiment 1 in Lazear et al. (2012) is a between-subjects design in which one group of subjects play a distribute \$10 dictator game and another group of subjects can sort out of the \$10 dictator game, and be paid \$10, or sort in and play the distribute \$10 dictator game. The extended game with the outside option is modeled as above with the MMA model using the unambiguous (10,0) endowment provided by the outside option. The distribute \$10 dictator game without outside options is a commonly used protocol for dictator games in which neither the dictator nor the recipient has a clearly assigned property right. This form of dictator game protocol is widely viewed as appropriate for research on sharing behavior but it does have an ambiguous endowment, as explained by Hoffman et al. (1994) and Hoffman, McCabe, and Smith (1996).[34] Experiment 1 data are consistent with predictions from the MMA model which follow from interpreting the 10 available for distribution as endowments to the dictator and recipient of $(10 - z, z)$, with $z > 0$.

6. IMPLICATIONS OF MMA FOR PLAY IN STRATEGIC GAMES WITH CONTRACTIONS

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

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

(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 Property α or a stronger traditional assumption such as convex preferences.

6.1 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.[35] Property α and MMA have different implications regarding the effects of 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 Property α 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, Property α 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 Property α makes this prediction only for FMs who take in the moonlighting game. Yet it is important to note that this latter "prediction" results solely from the constraint that prevents taking in the investment game, not from agent preferences in and of themselves.

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

[35] In the standard moonlighting game, the first and second mover are each endowed with equal amounts of money. The first mover can either give money or take money from the second mover, where the maximum amount that can be given is the full endowment but the maximum amount that can be taken is part of the endowment. Money given is transformed by a multiplier greater than 1 but money taken is not transformed. After the second mover learns about the outcome, he/she can also give or take money from the first mover at some cost. The investment game differs primarily in that the first mover can only give and not take.

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 consists of options that are costly for the SM but can increase/decrease FM's monetary payoff: $M_s = M_s^+ \cup M_s^-$ where

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

Consider the SM's choice in M_s in the Moonlighting Game when the FM sends a non-negative amount. Consistent with observed behavior[36] (as well as Pareto efficiency), the amount returned will be from M_s^+.

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

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

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

Suppose that 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. The FM choice set is

---

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

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

Thus, $I = M^+ \subset M$.

Property α 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, (FM coordinate, SM coordinate) = $(e/2, e)$ 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 Property α and MMA imply that for any *positive* amount received the SM's choices in the moonlighting and investment games are identical. We distinguish between two types of FMs: the ones who send in the moonlighting game and the ones who take. For a FM who takes in the moonlighting game, by design of the two games the FM must send more in the investment game. For a FM who does not take in the moonlighting game, we have shown above that Property α 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 Property α, 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 sent, we find that the means of the amounts sent are 5.97 (IG) and 4 (MG) and significantly different (t-test: one-sided p-value= 0.013.[37] Therefore the FM data are consistent with the above implications of MMA but inconsistent with implications of Property α.

[37] If we look only at Send > 0, the mean figures are 7.35 (IG) and 4.84 (MG) and significantly different (t-test: one-sided p-value=0.002).

SM choices: Estimates (standard errors in parentheses) of censored regressions for SM choices at information sets with "FM not taking" (send≥ 0, N=78) are[38]

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

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

6.2 Carrot, Stick, and Carrot/Stick Games

Andreoni, Harbaugh and Vesterlund (2003) look at effects of rewards and punishments on cooperation by studying behavior in three games: the carrot game that offers incentives only in terms of rewards, the Stick game that allows only for negative incentives (punishment) and the carrot and stick game (CS) that offers players both types of incentives. The two single incentive games are natural contractions of the CS game. We argue that for any given positive amount received the SM's predicted choice is *the same* in the CS and carrot game. This is the prediction of Property α 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 CS 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 amount returned, r_s by the SM can be positive (carrot), negative (stick) or either (CS game) as returning a negative amount is not feasible in the carrot game whereas returning a positive amount is not feasible in the stick game. Regardless of the sign of the amount returned, the FM receives $5r_s$.

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

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

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

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

Let r_{cs} be the SM's choice in the CS game when the FM sends amount *s*. Property α 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. Property α requires that if the SM choice in the CS 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 6) find larger demand for rewards in the CS game than in the carrot game which is inconsistent with both Property α 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. Property α requires that if the SM's most preferred choice in the CS 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 5) report a result they characterize as "surprising" (pg. 898) that demand for punishment is larger in the CS game than in the stick game. This result is predicted by MMA but is inconsistent with Property α.

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

7. 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). But the designs of these experiments do not provide an empirical challenge to rational choice theory. We take this next step by designing an experiment that generates data to test the empirical implications of Property α that is central to the theory. We find data that are inconsistent with extant rational choice theory. Our experimental design and data suggest why, and how, rational choice theory needs to be extended to maintain consistency with our data patterns.

In this spirit, we propose moral reference points as features of feasible sets and a moral monotonicity axiom (MMA). 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.

Development of the MMA model was motivated by an initial objective of rationalizing otherwise-anomalous data from dictator games with giving and taking opportunities. The model, however, has more general applicability. We explain how it can rationalize data from other types of dictator games in the literature. More importantly, we explain how the model has implications for play of strategic games involving contractions of feasible sets that differ from implications of extant 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 briefly provided an explanation of how the theory our morel reference points is predictive of received findings in a range of economic games designed to elicit social and cooperation behaviors. In this manner, we view our results as having both positive and normative import. For empiricists and practitioners, the results herein provide an indication that moral costs can play an important role in welfare calculations and program evaluation.

REFERENCES

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

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.

Bourlès, Renaud, Yann Bramoullé, and Eduardo Perez-Richet. "Altruism in networks." *Econometrica* 85, no. 2 (2017): 675-689.

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

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

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

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

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

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

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

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

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

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

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

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

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

Cox, James C., Vjollca Sadiraj, and Ulrich Schmidt. "Paradoxes and Mechanisms for Choice under Risk." *Experimental Economics* 18, no. 2 (2015): 215-250.

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.

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. "Experimental Investigations of Errors in Decision Making under Risk." *European Economic Review* 39, no. 3-4 (1995): 633-640.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

Zermelo, Ernst. "Beweis, dass jede Menge wohlgeordnet werden kann". *Mathematische Annalen* 59, no.4 (1904): 514–16. doi:10.1007/BF01445300.

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.

TABLES

Table 1. Summary Statistics

		Moral Reference Point	Ave. Transfer[a] (st. dev.)	Ave. Final Payoffs[b]	Nobs
Inequality	Give	(23,3)	4.54 (2.96)	(22.46, 7.54)	61
	Take	(19,3)	4.19 (3.34)	(22.81, 7.19)	81
	Symmetric	(15,3)	6.12 (4.95)	(20.88, 9.12)	82
Equal	Give	(19,7)	2.65 (2.24)	(20.35, 9.65)	66
	Take	(15,7)	3.17 (2.88)	(19.83, 10.17)	58
	Symmetric	(11,7)	3.94 (3.52)	(19.06, 10.94)	62
Envy	Give	(15,11)	2.43 (1.65)	(16.57, 13.43)	67
	Take	(11,11)	2.06 (1.85)	(16.94, 13.06)	69
	Symmetric	(7,11)	2.64 (2.55)	(16.36, 13.64)	66

a. "Ave. Transfer" is the amount by which the average recipient's payoff chosen by dictators exceeds the recipient's minimum expectations payoff (standard deviations in parentheses).
b. Final payoffs, with dictator payoff first followed by recipient payoff.

Table 2. Consistency with Convexity, Property α and MMA

Environments	Data with Convex Preferences Predictions		Data with Property α Predictions	
	Convexity	MMA	Property α	MMA
Inequality	39.36 (94)	78.72$^{.017}$ (94)	45.68 (81)	75.31 (81)
Equal	46.24 (93)	81.72$^{.016}$ (93)	48.31 (89)	80.90$^{.052}$ (89)
Envy	54.46 (101)	78.22 (101)	55.56 (99)	77.78 (99)
All	46.88 (288)	79.51$^{.003}$ (288)	50.19 (269)	78.07$^{.097}$ (269)

Note: Entries are percentages of choices consistent with predictions by the model in a column. Number of subjects in brackets. Entries as superscripts are one-sided p-values (when <.1) for the Sign Test. To conduct the Sign Test, observations that are consistent with Property α are coded as 0, the ones that are consistent with MMA (but violate Property α) are coded as 1, whereas observations that violate both (MMA and Property α) are coded as -1.

Table 3. Tests for Effects of Recipient Moral Reference Dimension

Recipient's Final Payoff	r_1 =15		r_1 =19		r_1 =11	
r_2 **[+]**	0.674***	0.668***	0.415*	0.391*	0.330**	0.328**
	(0.187)	(0.186)	(0.215)	(0.221)	(0.155)	(0.151)
Constant	6.145***	6.955***	6.435***	5.616***	8.620***	9.341***
	(1.548)	(2.417)	(1.143)	(1.895)	(1.480)	(1.797)
Demographics	no	yes	no	yes	no	yes
Observations	207	207	147	147	131	131
Log-likelihood	-261.3	-258.3	-224.8	-221.4	-225.9	-219.4

Notes: Entries are Tobit estimated coefficients. MMA predicted sign in square brackets. Demographics include gender, race, GPA, religion, major and study year. Standard errors in parentheses. *** p<0.001, **p<0.05, p*<0.1

Table 4. Moral Reference Points and Transfers

Dep. Variable	Hurdle Model		Tobit Model	
Transfer	(1)	(2)	(1)	(2)
r_1 [-]	-0.058**	-0.055**	-0.098**	-0.104**
	(0.027)	(0.027)	(0.047)	(0.047)
r_2 [-]	-0.319***	-0.314***	-0.497***	-0.487***
	(0.047)	(0.047)	(0.091)	(0.090)
Demographics	no	yes	no	yes
Observations	612	612	612	612

Notes: MMA predicted sign in square brackets. Entries are average marginal effects (Hurdle Model) and coefficients (Tobit model). Standard errors (clustered at subject ID level) in parentheses. Demographics include gender, race, GPA, religion, major and study year. Low and upper bounds in regressions are 8 and 0. [b] *** p<0.01, ** p<0.05, * p<0.1.

FIGURES

Figure 1. Feasible Sets: [B, C] for Give or Take, [A, C] for Symmetric

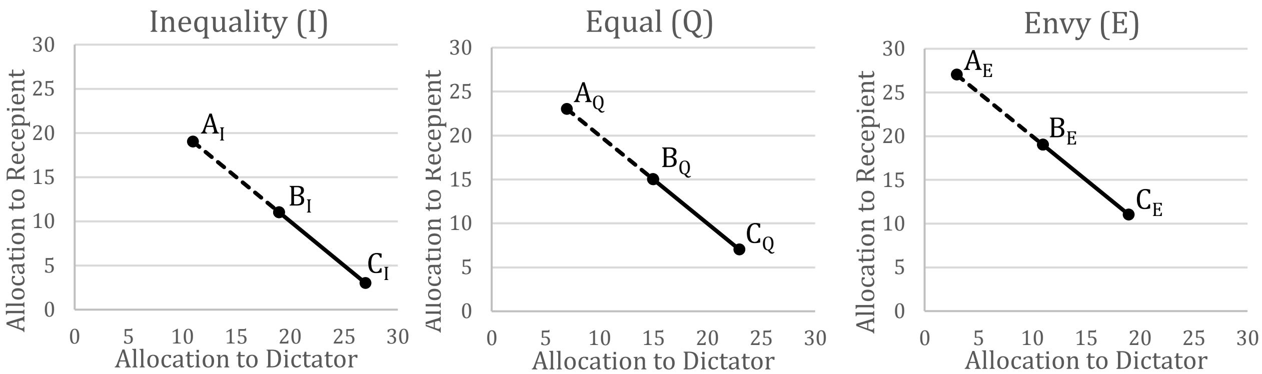
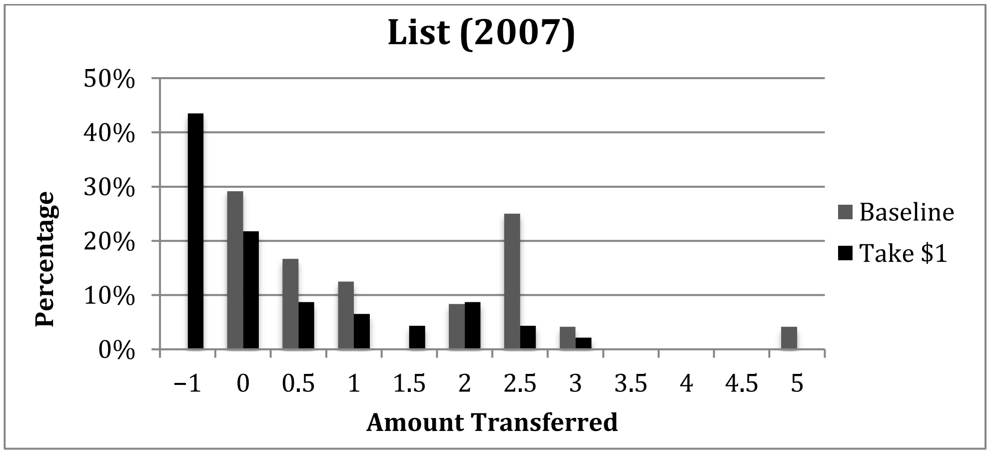
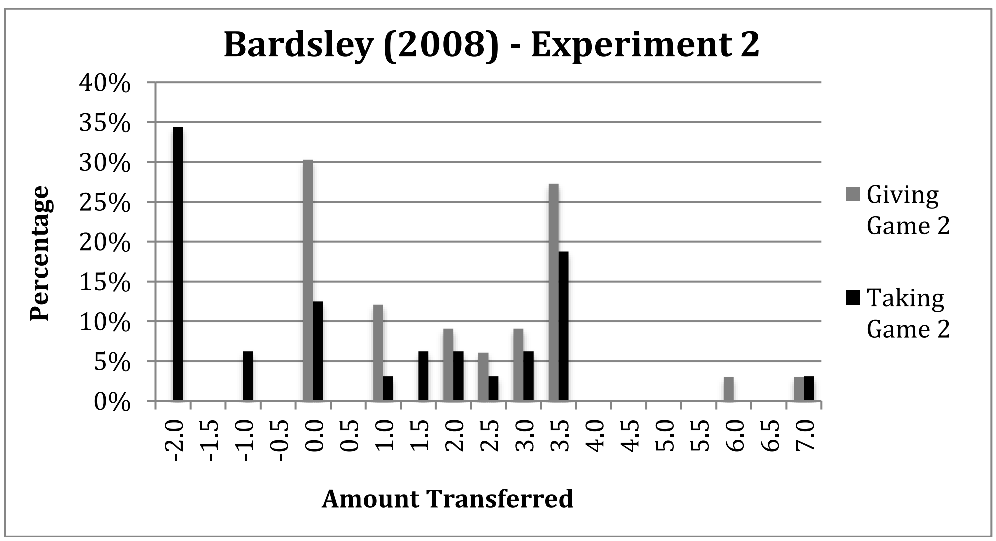


Figure 2. Histograms using Data from List (2007) and Bardsley (2008)





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 3. Example of Choice with non-Convex Preferences

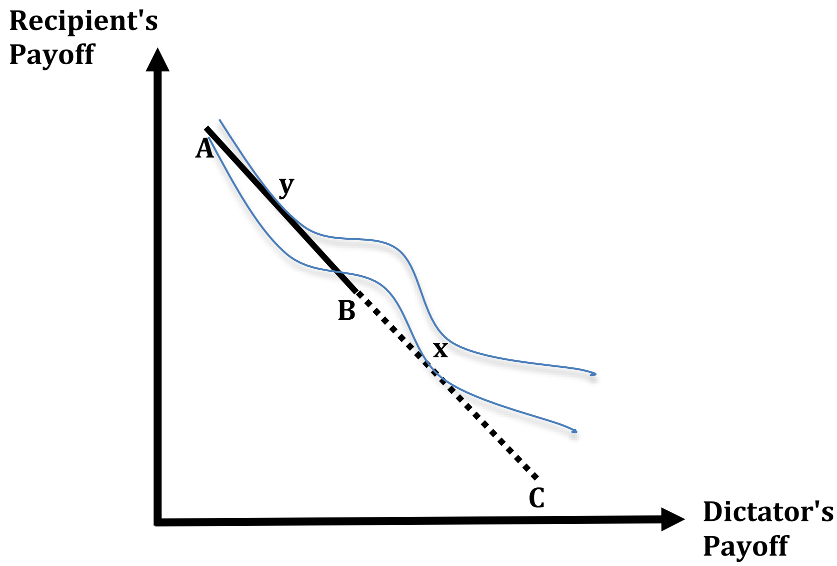
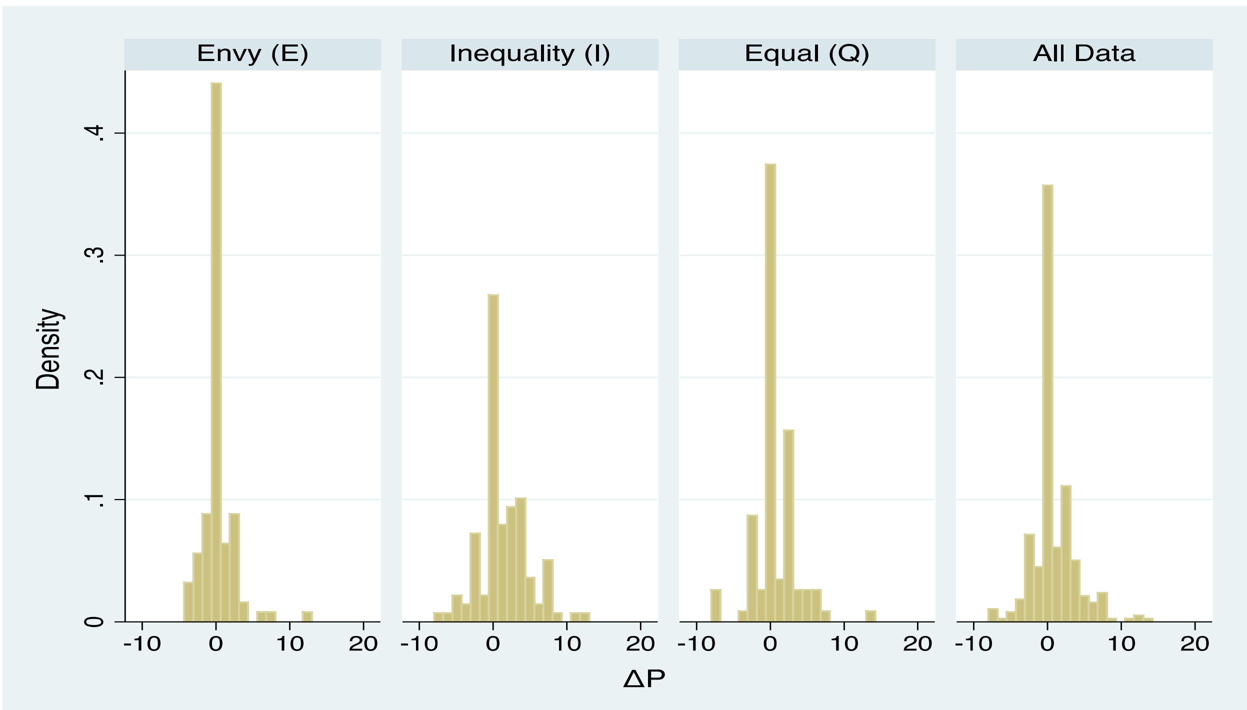


Figure 4. Histograms of ΔP (within subjects)



Notes: ΔP is constructed as follows. Each subject made two choices and for each choice the final own payoff was calculated. If the subject made a choice in Take and one in Give action sets then ΔP is payoff in Give minus payoff in Take. If the subject made one choice in Take and one choice in Symmetric actions sets then ΔP is payoff in Take minus payoff in Symmetric. Finally if the subject made one choice in Give and one choice in Symmetric actions sets then ΔP is payoff in Give minus payoff in Symmetric.

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

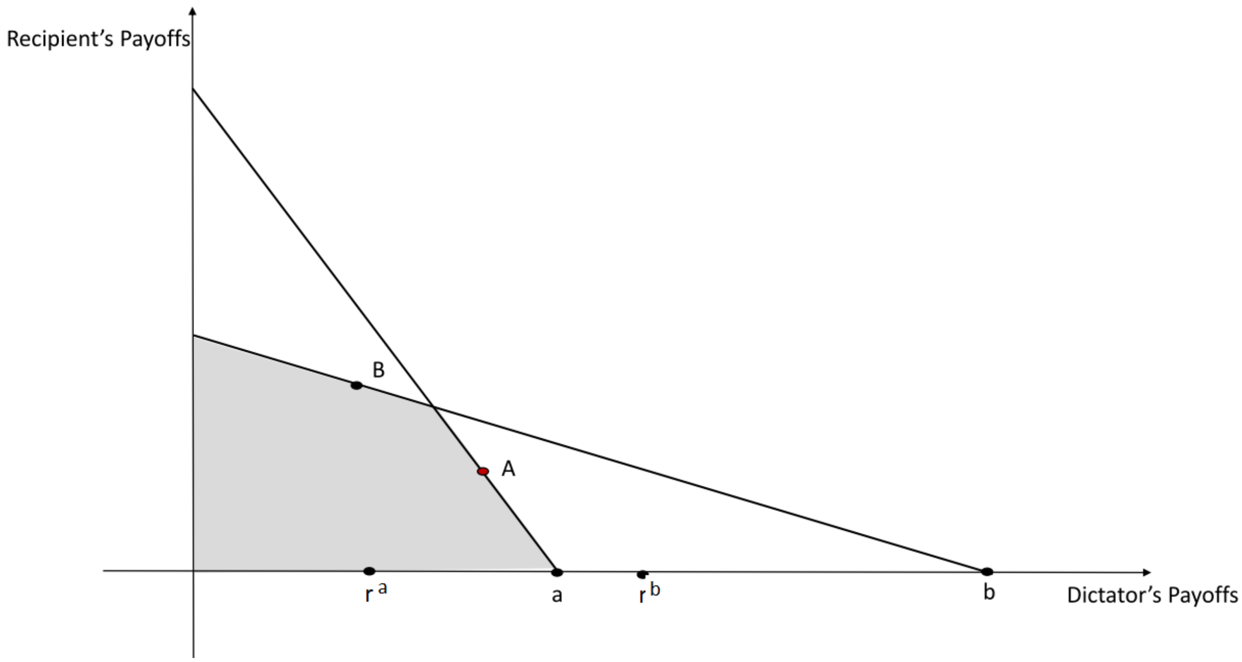
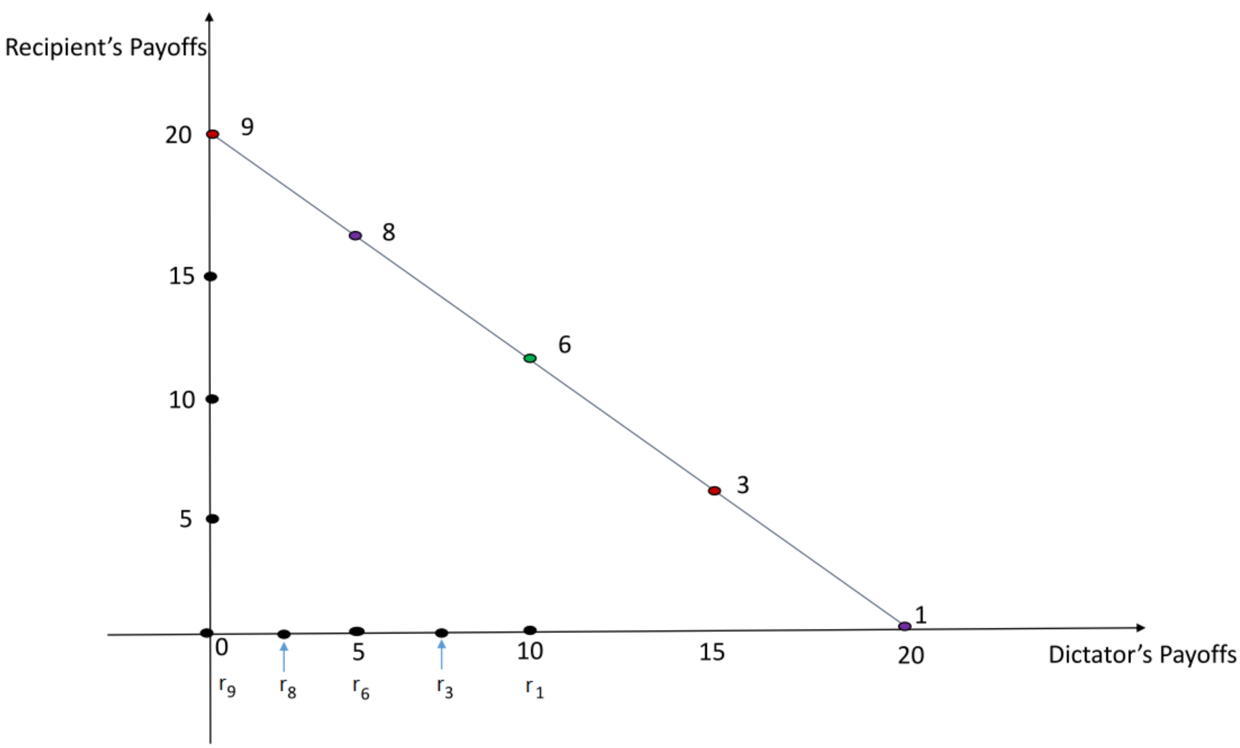


Figure 6. Endowments and Moral Reference Points for Korenok, et al. Treatments



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. Effect of Moral Reference Point on Transfers

"Transfer" is defined as the amount by which the recipient's payoff exceeds her minimum expectations payoff. In a Give treatment, the transfer is the amount the dictator gives to the recipient. In a Take treatment, the transfer is the amount the dictator does *not* taken from the recipient. In all treatments, the dictator makes a choice of an amount to give or take that we here represent by a transfer, $t \in T$, where *T*=[0,16] in the Symmetric version (Envy, Equal and Inequality) and *T*=[0,8] in the Give/Take scenarios (Envy, Equal and Inequality). The feasible set is

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

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

(Conventional) other-regarding preferences [39] Let $P^* = (m^*, y^*)$ be the dictator's most preferred allocation of \$30. Then when the dictator faces a set X that contains P^*, by Property α the dictator's choice of transfer t^o is such that $t^o + r_2 = y^*$. Thus, as r_2 increases, the most preferred transfer, t^o decreases for as long as the set X contains P^*, and it is 0 after that. Using utility representation terminology, let dictator's preferences be represented by some concave, *m*-increasing C[1] function $u(m, y)$. The dictator's decision problem is

[39] Such preferences include inequality aversion (Bolton and Ockenfels 2000, Fehr and Schmidt 1999), quasi-maxmin (Charness and Rabin 2002) or ego-centric altruism preferences (Cox and Sadiraj 2012).

$$max_{t\in T}u(m,y)=u\left(30-r_2-t,r_2+t\right)$$

By concavity of $u\left(.\right)$, the optimal t^o if from the interior of *T* solves $G\left(t,r\right)=\nabla u\cdot\left(-1,1\right)=0.$

Apply the implicit theorem to get[40] $sign\left(\frac{\partial t^o}{\partial r_2}\right)=sign\left(G_{r_2}\left(t^o,r\right)\right)<0$ and as $G\left(.\right)$ does not (directly) depend on r_1 , $\frac{\partial t^o}{\partial r_1}=0$. Thus we have the following hypothesis:

Hypothesis α : The optimal transfer, t^o decreases in r_2 but is not affected by r_1

MMA type of preferences Let dictator's choice satisfy MMA and suppose that dictator's choices can be recovered as a solution to the following maximization problem[41]

$$max_{t\in T}u(m-r_1,y-r_2)=u\left(30-\left(r_1+r_2\right)-t,t\right)$$

where $u(\cdot)$ is a concave C^1 function. Let t^r be the optimal transfer.

By MMA, dictator's optimal payoff, $m^*=30-\left(r_1+r_2\right)-t^r$ increases in r_1, i.e., $\frac{\partial m^*}{\partial r_1}=-1-\frac{\partial t^r}{\partial r_1}>0$. It follows that, (*) $\frac{\partial t^r}{\partial r_1}<0.$

Next, let $F\left(t,r\right)=\nabla u(m-r_1,y-r_2)\cdot\left(-1,1\right)$. By concavity, the optimal t^r (at the interior) solves $F\left(t^r,r\right)=0$. To show that $\frac{\partial t^r}{\partial r_2}<0$ use (*) and the implicit function theorem:

$$sign\left(\frac{\partial t^r}{\partial r_2}\right)=sign\left(\frac{\partial F}{\partial r_2}\right)=sign\left(\frac{\partial F}{\partial r_1}\right)=sign\left(\frac{\partial t^r}{\partial r_1}\right).$$

Thus we have the following hypothesis:

Hypothesis MMA: The optimal transfer, t^r decreases in r_2 as well as in r_1

---

[40] By concavity of $u(\cdot)$, $G_{r_2}\left(t^o,r_2\right)=u_{11}-2u_{12}+u_{22}=Q\left(-1,1\right)<0.$.

[41] One way to think about a dictator who is "socially" cautious is that he can claim social credits only for the transfer part, *t* rather than all recipient's payoff, r_2+t (as the recipient gets r_2 no matter what by the experimenter). If so then rather than the distribution of 30, the dictator's problem is the distribution of $(30-(r_2+r_1))$ between oneself and the other person.