Revealed Altruism

By James C. Cox, Daniel Friedman, and Vjollca Sadiraj

Forthcoming in Econometrica

REVEALED ALTRUISM

BY JAMES C. COX, DANIEL FRIEDMAN, AND VJOLLCA SADIRAJ

ABSTRACT. This paper develops a nonparametric theory of preferences over one's own and others' monetary payoffs. We introduce "more altruistic than" (**MAT**), a partial ordering over such preferences, and interpret it with known parametric models. We also introduce and illustrate "more generous than" (**MGT**), a partial ordering over opportunity sets. Several recent studies focus on two player extensive form games of complete information in which the first mover (FM) chooses a more or less generous opportunity set for the second mover (SM). Here reciprocity can be formalized as the assertion that an **MGT** choice by the FM will elicit **MAT** preferences in the SM. A further assertion is that the effect on preferences is stronger for acts of commission by FM than for acts of omission. We state and prove propositions on the observable consequences of these assertions. Finally, empirical support for the propositions is found in existing data from Investment and Dictator games, the Carrot and Stick game, and the Stackelberg duopoly game and in new data from Stackelberg mini-games.

KEYWORDS: Neoclassical Preferences, Social Preferences, Convexity, Reciprocity, Experiments.

1. INTRODUCTION

What are the contents of preferences? People surely care about their own material well-being, e.g., as proxied by income. In some contexts people also may care about others' well-being. Abstract theory and common sense have long recognized that possibility but until recently it has been neglected in applied work. Evidence from the laboratory and field (as surveyed in Fehr and Gächter (2000), for example) has begun to persuade economists to develop specific models of how and when a person's preferences depend on others' material payoffs (Sobel (2005)).

[0]For helpful comments, we thank James Andreoni, Geert Dhaene, Steven Gjerstad, Stephen Leider, Joel Sobel, Stefan Traub, and Frans van Winden as well as participants in the International Meeting of the Economic Science Association (ESA) 2004, the North American Regional ESA Meeting 2004, and at Economics Department seminars at UCSC, Harvard and University College London. The final revision is much improved due to the suggestions of three anonymous referees and Associate Editor David Levine. Financial support was provided by the National Science Foundation (grant numbers IIS-0630805 and IIS-0527770).

Andreoni and Miller (2002) report "dictator" experiments in which a human subject decides on an allocation for himself and for some anonymous other subject while facing a linear budget constraint. Their analysis confirms consistency with the generalized axiom of revealed preference (GARP) for a large majority of subjects. They conclude that altruism can be modeled using neoclassical preference theory (Hicks (1939), Samuelson (1947)).

In this paper we take three further steps down the same path. First, we analyze non-linear opportunity sets. Such sets allow a player to reveal more about the tradeoff between her own and another's income, e.g., whether her indifference curves have positive or negative slope, and whether they are linear or strictly convex. Second, we give another player an initial move that can be more or less generous. This allows us to distinguish conditional altruism—positive and negative reciprocity—from unconditional altruism. It also allows us to clarify the observable consequences of convex preferences and of reciprocal preferences. Third, we distinguish active from passive initial moves; i.e., we distinguish among acts of omission, acts of commission, and absence of opportunity to act, and examine their impacts on reciprocity.

Our goal is to develop an approach to reciprocity firmly grounded in neoclassical preference and demand theory.[1] By contrast, much of the existing literature on social preferences either ignores reciprocity motives or grounds them in psychological game theory. Our focus is on how players' choices respond to observable events and opportunities, rather than to their beliefs about other players' intentions or types.

Section 2 begins by developing representations of preferences over own and others' income, and formalizes the idea that one preference ordering is "more altruistic than" (**MAT**) another. It allows for the possibility of negative regard for the other's income; in this case **MAT** really means "less malevolent than." Special cases include the main parametric models of other-regarding preferences that have appeared in the literature.

Section 3 introduces opportunities and formalizes the idea that one opportunity set can be more generous than (**MGT**) another. It explains that **MGT** is a partial ordering over standard budget sets and is a complete ordering over opportunity sets in several two player games, including the well-known Investment, Dictator, and Stackelberg duopoly games. An appendix demonstrates

[1]Cox, Friedman and Gjerstad (2007) takes a similar perspective, but it imposes a tight parametric structure (CES) on preferences and reports structural estimates from various existing data sets. Here we seek general results attributable to general properties such as convexity and reciprocity, and we test the results directly on new as well as existing data.

MGT orderings of opportunity sets for several other games in the literature on social preferences.

Section 4 formalizes reciprocity. Axiom R asserts that more generous choices by a first mover induce more altruistic preferences in a second mover. An interpretation (advocated in Cox, Friedman, and Gjerstad (2007)) is that preferences are emotional state-dependent, and the first mover's generosity induces a more benevolent (or less malevolent) emotional state in the second mover. Axiom S asserts that the reciprocity effect is stronger following an act of commission (upsetting the status-quo) than following an act of omission (upholding the status-quo), and that the effect is weaker when the first mover is unable to alter the status quo.

Section 5 presents three general theoretical propositions on the consequences of convex preferences. Among other things, these propositions extend standard results on revealed preference theory and show how easy it is in empirical work to conflate the separate effects of convexity and reciprocity.

Sections 6 - 9 bring revealed altruism theory to four data sets. Proposition 4 derives testable predictions for Investment and Dictator games. Together, these two games provide diagnostic data for both Axiom R and Axiom S. Propositions 5 and 6 derive testable predictions for Carrot and Stick games and for Stackelberg duopoly games. The duopoly games are especially useful because the Follower's opportunity sets are **MGT**-ordered and have a parabolic shape that enables the Follower to reveal a wide range of positive and negative tradeoffs between his own income and Leader's income. Proposition 7 obtains predictions for a new variant game, called the Stackelberg mini-game, in which the Leader has only two alternative output choices, one of which is clearly more generous than the other. This game provides diagnostic data for discriminating between the effects of convexity and reciprocity.

Within the limitations of the data, the test results are consistent with predictions. Following a concluding discussion, Appendix A collects all formal proofs and other mathematical details. Instructions to subjects in the Stackelberg mini-game appear in Appendix B.

2. Preferences

Let $Y = (Y_1, Y_2, ..., Y_N) \in \mathfrak{R}^N_+$ represent the payoff vector in a game that pays each of $N \geq 2$ players a non-negative income. Admissible preferences for each player i are smooth and convex orderings on the positive orthant $\mathfrak{R}^N_+$ that are strictly increasing in own income Y_i. The set of all admissible

preferences is denoted $\mathfrak{P}$. Any particular preference $\mathcal{P} \in \mathfrak{P}$ can be represented by a smooth utility function $u : \Re_+^N \to \Re$ with positive i^{th} partial derivative $\partial u/\partial Y_i = u_{Y_i} > 0$. The other first partial derivatives are zero for standard selfish preferences, but we allow for the possibility that they are positive in some regions (where the agent is "benevolent") and negative in others (where she is "malevolent").

We shall focus on two-player extensive form games of complete information, and to streamline notation we shall denote own ("my") income by $Y_i = m$ and the other player's ("your") income by $Y_{-i} = y$. Thus preferences are defined on the positive quadrant $\Re_+^2 = \{(m, y) : m, y \geq 0\}$. The marginal rate of substitution $\text{MRS}(m, y) = u_m/u_y$ is not well defined at points where the agent is selfish; it diverges to $+\infty$ and back from $-\infty$ when we pass from slight benevolence to slight malevolence. Therefore it is convenient to work with willingness to pay, WTP= 1/MRS, the amount of own income the agent is willing to give up in order to increase the other agent's income by a unit; it moves from slightly positive through zero to slightly negative when the agent goes from slight benevolence to slight malevolence. Note that WTP= u_y/u_m is intrinsic, independent of the particular utility function u chosen to represent the given preferences.

What sort of factors might affect w =WTP? Of course, for admissible preferences the sign of w is the same as the sign of the partial derivative u_y. Convexity tells us more: w increases as one moves southward along an indifference curve. That is, my benevolence increases (or malevolence decreases) as your income decreases along an indifference curve. This principle is quite intuitive, and sometimes it is useful to strengthen it as follows. We say that admissible preferences have the *increasing benevolence* (IB) property if $w_m \geq 0$. Occasionally we refer to the related property $w_y \leq 0$. Appendix A.1 shows how convexity, increasing benevolence, and homotheticity are related to each other and to the slope and curvature of indifference curves.

We are now prepared to formalize the idea that one preference ordering on $\Re_+^2$ is more altruistic than another. Two different preference orderings $\mathcal{A}, \mathcal{B} \in \mathfrak{P}$ over income allocation vectors might represent the preferences of two different players, or might represent the preferences of the same player in two different situations.

Definition 1. For a given domain D$\subset \Re_+^2$ we say that $\mathcal{A}$ **MAT** $\mathcal{B}$ on D if $\text{WTP}_{\mathcal{A}}(m, y) \geq \text{WTP}_{\mathcal{B}}(m, y)$, for all $(m, y) \in$D.

The idea is straightforward. Like the single crossing property in a different context, **MAT** induces a partial ordering on preferences over own and others' income. In the benevolence case, $\mathcal{A}$ **MAT** $\mathcal{B}$ means that $\mathcal{A}$ has shallower indifference curves than $\mathcal{B}$ in (m, y) space, so $\mathcal{A}$ indicates a willingness to pay more m for a unit increase in y than does $\mathcal{B}$. In the malevolence case, WTP is less negative for $\mathcal{A}$, so it indicates a lesser willingness to pay for a unit decrease in y.

Appendix A.2 verifies that **MAT** is a partial ordering on $\mathfrak{P}$. When no particular domain D is indicated, the **MAT** ordering is understood to refer to the entire positive orthant D$= \Re^2_+$.

Four examples illustrate how **MAT** is incorporated into existing parametric models.

Example 2.1. Linear Inequality-averse Preferences (for $N = 2$ only; Fehr and Schmidt (1999)). Let preferences $\mathcal{J} = \mathcal{A}, \mathcal{B}$ be represented by $u_{\mathcal{J}}(m, y) = (1 + \theta_{\mathcal{J}})m - \theta_{\mathcal{J}} y$, where

$$\begin{aligned} \theta_{\mathcal{J}} &= \alpha_{\mathcal{J}}, \ \texttt{if } m < y \\ &= -\beta_{\mathcal{J}}, \ \texttt{if } m \geq y, \end{aligned}$$

with $\beta_{\mathcal{J}} \leq \alpha_{\mathcal{J}}$ and $0 < \beta_{\mathcal{J}} < 1$. Straightforwardly, $\mathcal{A}$ **MAT** $\mathcal{B}$ if and only if $\theta_{\mathcal{A}} \leq \theta_{\mathcal{B}}$.

Example 2.2. Nonlinear Inequality-averse Preferences (for $N = 2$, Bolton and Ockenfels (2000)). Let preferences $\mathcal{J} = \mathcal{A}, \mathcal{B}$ be represented by $u_{\mathcal{J}}(m, y) = \nu_{\mathcal{J}}(m, \sigma)$, where

$$\begin{aligned} \sigma &= m/(m + y), \ \texttt{if } m + y > 0 \\ &= 1/2, \ \texttt{if } m + y = 0. \end{aligned}$$

It can be easily verified that $\mathcal{A}$ **MAT** $\mathcal{B}$ if and only if $\nu_{\mathcal{A}1}/\nu_{\mathcal{A}2} \leq \nu_{\mathcal{B}1}/\nu_{\mathcal{B}2}$.

Example 2.3. Quasi-maximin Preferences (for $N = 2$, Charness and Rabin (2002)). Let preferences $\mathcal{J} = \mathcal{A}, \mathcal{B}$ be represented by

$$\begin{aligned} u_{\mathcal{J}}(m, y) &= m + \gamma_{\mathcal{J}}(1 - \delta_{\mathcal{J}})y, \ \texttt{if } m < y \\ &= (1 - \delta_{\mathcal{J}}\gamma_{\mathcal{J}})m + \gamma_{\mathcal{J}} y, \ \texttt{if } m \geq y, \end{aligned}$$

and $\gamma_{\mathcal{J}} \in [0, 1]$, $\delta_{\mathcal{J}} \in (0, 1)$. It is straightforward (although a bit tedious) to verify that $\mathcal{A}$ **MAT** $\mathcal{B}$ if and only if

$$\gamma_{\mathcal{A}} \geq \gamma_{\mathcal{B}} \max \left\{ \frac{1}{1 + (\delta_{\mathcal{A}} - \delta_{\mathcal{B}})\gamma_{\mathcal{B}}}, \frac{1 - \delta_{\mathcal{B}}}{1 - \delta_{\mathcal{A}}} \right\}.$$

Example 2.4. Egocentric Altruism (CES) Preferences (Cox and Sadiraj (2007)). Let preferences $\mathcal{J}$=$\mathcal{A},\mathcal{B}$ be represented by

$$\begin{aligned} u_{\mathcal{J}}(m,y) &= \frac{1}{\alpha}(m^{\alpha}+\theta_{\mathcal{J}}y^{\alpha}),\ \texttt{if}\ \alpha\in(-\infty,1)\diagdown\{0\} \\ &= my^{\theta_{\mathcal{J}}},\ \texttt{if}\ \alpha=0. \end{aligned}$$

If $0<\theta_{\mathcal{B}}\leq\theta_{\mathcal{A}}$ then $\mathcal{A}$ **MAT** $\mathcal{B}$. Verification is straightforward: $WTP_{\mathcal{J}}=\theta_{\mathcal{J}}(m/y)^{1-\alpha}$, $\mathcal{J}=\mathcal{A},\mathcal{B}$ imply $WPT_{\mathcal{A}}/WTP_{\mathcal{B}}=\theta_{\mathcal{A}}/\theta_{\mathcal{B}}\geq 1$. "Egocentricity" means that $u_{\mathcal{J}}(x+\epsilon,x-\epsilon)>u_{\mathcal{J}}(x-\epsilon,x+\epsilon)$ for any $\epsilon\in(0,x)$ which implies $WTP(m,m)\leq 1$.

Much of the theoretical literature on social preferences relies on special assumptions that may appear to be departures from neoclassical preference theory (Hicks (1939), Samuelson (1947)). The preceding examples help clarify the issues. All four are examples of convex preferences, and (except for the nonlinear inequality aversion model) they are also homothetic. The inequality aversion models incorporate a very specific inconsistency with the neoclassical assumption of positive monotonicity: my marginal utility for your income reverses sign on the 45 degree line. A preference for efficiency (i.e., for a larger income sum) is consistent with a limiting case of the quasi-maximin model, or with admissible preferences with $WTP=1$. We shall now see that for more general preferences, the efficiency of choices depends on the shape of the opportunity set.

3. Opportunities

Define an opportunity set F (or synonymously, a feasible set or budget set) as a convex compact subset of $\Re^2_+$. It is convenient and harmless (given preferences monotone in own income m) to assume free disposal for own income, i.e., if $(m,y)\in F$ then $(am,y)\in F$ for all $a\in[0,1]$. Thus an opportunity set F is the convex hull of two lines: (a) its projection $Y_F=\{y\geq 0:\exists m\geq 0\ \texttt{s.t.}\ (m,y)\in F\}$ on the y-axis, and (b) its Eastern boundary $\partial_E F=\{(m,y)\in F:\forall x>m,(x,y)\notin F\}$.

Since F is convex, each boundary point has a supporting hyperplane (i.e., tangent line) defined by an inward-pointing normal vector, and F is contained in its closed positive halfspace; see for example Rockafellar (1970, p. 100). At some boundary points (informally called corners or kinks) the supporting hyperplane is not unique; examples will be noted later. At the other (regular) boundary points there is a smooth function f whose zero isoquant defines the boundary locally. We often need to work near vertical tangents, so rather

than the usual marginal rate of transformation (MRT) we use the need to pay, $\text{NTP}(m,y) = 1/\text{MRT}(m,y) = f_y/f_m$ evaluated at a regular point $(m,y) \in \partial_E F$. Again NTP is intrinsic, independent of the choice f used to represent the boundary segment.

We seek an objective definition of one opportunity set G being more generous to me than another opportunity set F. There is an obvious necessary condition: that G allows me to achieve higher income than does F. Since my preferences are monotone in own income, I clearly benefit when you allow me to increase it. For some purposes it is helpful to impose a second condition, that you don't increase your own potential income far more than mine. If you do, I might regard your move as self-serving and not especially generous.

These intuitions are captured in conditions (a) and (b) below, using the following notation. Let $y_F^* = \sup Y_F$ denote your maximum feasible income and let $m_F^* = \sup\{m : \exists y \geq 0 \text{ s.t. } (m,y) \in F\}$ denote my maximum feasible income in an opportunity set F.

Definition 2. Opportunity set $G \subset \Re_+^2$ is more generous than opportunity set $F \subset \Re_+^2$ if (a) $m_G^* - m_F^* \geq 0$ and (b) $m_G^* - m_F^* \geq y_G^* - y_F^*$. In this case we write G **MGT** F.

MGT is a partial ordering over opportunity sets, as noted in Appendix A.3. Condition (a) seems compelling because it springs directly from the most basic intuitions about generosity, but one can imagine plausible variants on condition (b). To understand its role, consider an alternative definition of **MGT**, call it **MGT** Light, that includes only condition (a). It turns out that **MGT** Light has the same implications as **MGT** for ten of the twelve prominent examples of opportunity sets from the social preferences literature discussed in this section, section 9 and Appendix A.5. We begin with a very prominent example where condition (b) does matter.

Example 3.1. Standard budget set. Let $F = \left\{(m,y) \in \Re_+^2 : m + py \leq I\right\}$ for given $p, I > 0$. Then the Eastern boundary $\partial_E F$ is the budget line $\left\{(m,y) \in \Re_+^2 : m + py = I\right\}$, as shown by the solid line in Figure 1. The NTP is p along $\partial_E F$. Clearly $m_F^* = I$ and $y_F^* = I/p$. To illustrate the **MGT** ordering, let F be determined by I_F and p_F and G by I_G and p_G. Part a of the definition is simply $I_G \geq I_F$. But part b requires $I_G - I_F \geq I_G/p_G - I_F/p_F$. For example, if $I_G = 1.1 I_F$ while $p_G = p_F/100$ so $y_G^* = 110 y_F^*$, as shown by the dashed line in Figure 1, then you have not clearly revealed generosity towards me by choosing G over F, since you are serving your own material interests far

more than mine. Your choice would more clearly reveal generosity if G (and F) were also consistent with part b.

Example 3.2. Investment game (Berg, Dickhaut, and McCabe (1995)). In this two player sequential move game, the First Mover (FM) and the Second Mover (SM) each have an initial endowment of $I \geq 1$. The FM sends an amount $s \in [0, I]$ to SM, who receives ks. Then the SM returns an amount $r \in [0, ks]$ to the FM, resulting in payoffs $m = I+ks-r$ for SM and $y = I-s+r$ for FM. The FM's choice of s selects the SM's opportunity set F_s with Eastern boundary $\left\{(m, y) \in \Re^2_+ : m + y = 2I + (k-1)s, m \in [I, I + ks]\right\}$ with NTP = 1. Figure 2 shows F_s for $s = 3$ and 9 when $I = 10$ and $k = 3$. In the figure, one sees that (a) $m^*_{F_9} = 37 > 19 = m^*_{F_3}$ and (b) $y^*_{F_9} - y^*_{F_3} = 28 - 16 = 12 < 18 = 37 - 19 = m^*_{F_9} - m^*_{F_3}$, so F_9 **MGT** F_3. More generally, it is straightforward to check that $s > s' \in [0, I]$ implies for $k \geq 2$ that F_s **MGT** $F_{s'}$, i.e., sending a larger amount is indeed more generous.

Example 3.3. Carrot and/or Stick Games (Andreoni, Harbaugh, and Vesterlund (2003)). In each of the games, the FM has an initial endowment of 240 and the SM has an initial endowment of 0. The FM sends an amount $s \in [40, 240]$ to SM, who receives s. The SM then returns an amount r which is multiplied by 5 for the FM, resulting in payoffs $m = s - |r|$ for SM and $y = 240 - s + 5r$ for FM.

The games differ only on the sign restrictions placed on r. In the Stick game, the SM can punish the FM at a personal cost by "returning" nonpositive amounts r that do not make either person's payoff negative. The FM's choice s induces an **MGT**-ordering on the SM opportunity sets F_s. Part a of the definition is satisfied because $m^*_{F_s} = s$ and part b is satisfied because $y^*_{F_s} = 240 - s$. For $F = F_s$ and $G = F_{s'}$ with $s < s^{'}$, we have $y^*_G - y^*_F = -(s^{'} - s) < 0 < s^{'} - s = m^*_G - m^*_F$.

In the Carrot game, the SM's choice must be non-negative, $r \in [0, s]$. Here the FM's choice s does not induce an **MGT**-ordering on the SM opportunity sets F_s. Of course, $m^*_{F_s} = s$ still ensures that part a of MGT is satisfied and thus the opportunity sets are **MGT** Light ordered. However, $y^*_{F_s} = 240 - s + 5s = 240 + 4s$. For $F = F_s$ and $G = F_{s'}$ with $s < s^{'}$, we have $y^*_G - y^*_F = 4(s^{'} - s) > s^{'} - s = m^*_G - m^*_F$, contradicting part b of the **MGT** definition.

The Carrot-Stick game drops the sign restrictions on the SM's choice: here the positive or negative amounts returned r cannot make either person's payoff

negative. As in the Carrot game, the SM opportunity sets are not **MGT**-ordered because part b is not satisfied (though they are ordered by **MGT** Light).

Example 3.4. Stackelberg duopoly game (e.g., Varian (1992, p. 295-298)). Consider a duopoly with zero fixed cost, constant and equal marginal cost, and nontrivial linear demand. Without further loss of generality one can normalize so that the profit margin (price minus marginal cost) is $M = T - q_L - q_F$, where $q_L \in [0, T]$ is the Leader's output choice and $q_F \in [0, T - q_L]$ is the Follower's output to be chosen. Thus payoffs are $m = Mq_F$ and $y = Mq_L$. The Follower's opportunity set $F(q_L)$ has as its Eastern boundary a parabolic arc opening towards the y-axis, as shown in Figure 3 for $T = 24$ and $q^L = 6, 8$ and 11. Unlike the earlier examples, the NTP varies smoothly from negative to positive values as one moves northward along the boundary. These opportunity sets are **MGT** ordered by the Leader's output choice; see Section A.4 of the Appendix for a verification and for explicit formulas for NTP.

These four examples are far from exhaustive. Section A.5 of the Appendix demonstrates natural **MGT** orderings of opportunity sets in many games prominent in the social preferences literature, including the Ultimatum game (Güth, Schmittberger, and Schwarze (1982), the Ultimatum mini-game (Gale, Binmore, and Samuelson (1995); see also Falk, Fehr, and Fischbacher (2003)), the Sequential public goods game with two players (Varian (1994)), the Gift exchange labor market (Fehr, Kirchsteiger, and Riedl (1993)), the Moonlighting game (Abbink, Irlenbusch, and Renner (2000)), the Power to Take game (Bosman and van Winden (2002)), and the Ring test (Liebrand (1984); see also Sonnemans, van Dijk and van Winden (2005)).

4. Reciprocity

Reciprocity is key to our analysis. We examine it from the perspective of neoclassical preference theory, stressing observables. Thus positive reciprocity reveals itself via preferences for altruistic actions that benefit someone else, at one's own material cost, *because* that person's behavior was generous. Similarly, negative reciprocity reveals itself via preferences for actions that harm someone else, at one's own material cost, *because* that person's behavior was harmful to oneself. Our reciprocity axiom states that more generous choices by one player induce more altruistic preferences in a second player; by the same token, less generous choices by one induce less altruistic preferences in the other.

To formalize, consider a two person extensive form game of complete information in which the first mover chooses an opportunity set $C \in \mathcal{C}$, and the second mover chooses the payoff vector $(m, y) \in C$. Initially, the second mover knows the collection $\mathcal{C}$ of possible opportunity sets. Prior to her choice of payoffs, she learns the actual opportunity set $C \in \mathcal{C}$, and acquires preferences $\mathcal{A}_C$. Reciprocity is captured in

> **Axiom R:** *Let the first mover choose the actual opportunity set for the second mover from the collection $\mathcal{C}$. If $F, G \in \mathcal{C}$ and G* **MGT** *F, then $\mathcal{A}_G$* **MAT** *$\mathcal{A}_F$.*

There is a traditional distinction between sins of commission (active imposition of harm) and sins of omission (failure to prevent harm). By analogy, one can draw a distinction between "virtues" of commission and omission. Another person's benevolent or malevolent intentions are more clearly revealed by an action that overturns the status quo than by inaction. Of course, sometimes there is no choice possible; the status quo cannot be altered. Intuitively, the second mover will respond more strongly to generous (or ungenerous) choices that overturn the status quo than to those that uphold it, or that involve no real choice by the first mover.[2] Compared to no choice, upholding the status quo should provoke the stronger response, at least when the status quo is the best or worst possible opportunity.

To formalize the intuition, suppose that the collection of opportunity sets $\mathcal{C}$ contains at least two elements, and one of them, C^*, is the status quo. Let $\mathcal{A}_{C^*}$ and $\mathcal{A}_{C^c}$ respectively denote the second mover's acquired preferences when the first mover's chosen opportunity set C is the status quo and when it differs from the status quo. On the other hand, when $\mathcal{C}$ is a singleton, then the first mover has no choice and we write $\mathcal{C} = \{C^o\}$ with corresponding second mover preferences $\mathcal{A}_{C^o}$.

> **Axiom S:** *Let the first mover choose the actual opportunity set for the second mover from the collection $\mathcal{C}$. If the status quo is either F or G and G* **MGT** *F then*
>
> (1) $\mathcal{A}_{G^c}$ **MAT** $\mathcal{A}_{G^*}, \mathcal{A}_{G^o}$ and $\mathcal{A}_{F^*}$, $\mathcal{A}_{F^o}$ **MAT** $\mathcal{A}_{F^c}$,
> (2) $\mathcal{A}_{G^*}$ **MAT** $\mathcal{A}_{G^o}$ if G **MGT** C for all $C \in \mathcal{C}$, and $\mathcal{A}_{F^o}$ **MAT** $\mathcal{A}_{F^*}$ if C **MGT** F for all $C \in \mathcal{C}$.

Part 1 of Axiom S says that the effect of Axiom R is stronger when a generous (or ungenerous) act upsets the status quo than when the same act

[2]This intuition goes back at least to Adam Smith's *Theory of Moral Sentiments*, <1759> (1976, p. 181).

merely upholds the status quo (or is forced). Part 2 compares the impact of upholding the status quo to forced acts. It says that the effect of Axiom R is stronger for upholding the status quo, at least when that is the most (or least) generous of the options available to the first mover.

We will say that either axiom *holds strictly* when the inequalities in the **MAT** and the **MGT** part a definitions are both strict.

It should be emphasized that the recent preference models noted in Examples 2.1 - 2.4 have no room for Axioms R and S. In those models preferences are assumed fixed, unaffected by more or less generous opportunity sets chosen by the first mover. Actual choices by a first mover are not central even in the "reciprocity" models of Charness and Rabin (2002, Appendix), Falk and Fischbacher (2006), and Dufwenberg and Kirchsteiger (2004). Those models focus on higher-order beliefs regarding other players' intentions (or, in Levine (1998), regarding other players' types). Cox, Friedman, and Gjerstad (2007) implicitly consider Axiom R, but only within the particular parametric family of CES utility functions noted in Example 2.4.

5. Choice

As in neoclassical theory, our maintained assumption is that the player always chooses a most preferred point in his opportunity set F. By convexity such points must form a connected subset of F. If either preferences $\mathcal{A}$ or opportunities F are strictly convex then that subset is a singleton, i.e., there is a unique choice $(m_{\mathcal{A}}, y_{\mathcal{A}}) \in F$. In this case all points in $F \setminus \{(m_{\mathcal{A}}, y_{\mathcal{A}})\}$ are revealed to be on lower $\mathcal{A}$-indifference curves than $(m_{\mathcal{A}}, y_{\mathcal{A}})$.

Not all elements of F are candidates for choice in our set up. The first result is that, due to strict monotonicity in own payoff m, only points on the Eastern boundary will be chosen, since they have larger own payoff.

Proposition 1. *Let $(m_{\mathcal{A}}, y_{\mathcal{A}})$ be an $\mathcal{A}-$chosen point in F. Then $(m_{\mathcal{A}}, y_{\mathcal{A}}) \in \partial_E F$. The choice is unique if either the preferences $\mathcal{A}$ or the opportunity set F is strictly convex.*

All proofs are collected in Appendix A.

The next result shows that, as admissible preferences go from maximally malevolent through neutral to maximally benevolent under the **MAT** ordering, the player's choices trace out the entire Eastern boundary of the opportunity set. The proposition refers to the North point $N_F = (m, y_F^*) \in \partial_E F$ and the South point S_F, the point in the Eastern boundary with smallest y-component.

Proposition 2. *Suppose that either preferences $\mathcal{A}$ and $\mathcal{B}$, or the opportunity set F, are strictly convex. Let $(m_{\mathcal{A}}, y_{\mathcal{A}})$ and $(m_{\mathcal{B}}, y_{\mathcal{B}})$ be the points in F chosen when preferences are respectively $\mathcal{A}$ and $\mathcal{B}$. Then*

(1) *$\mathcal{B}$* **MAT** *$\mathcal{A}$ implies $y_{\mathcal{B}} \geq y_{\mathcal{A}}$.*
(2) *If $(m, y) \in \partial_E F$ and $y_{\mathcal{B}} \geq y \geq y_{\mathcal{A}}$, then there are preferences $\mathcal{P}$ with $\mathcal{B}$* **MAT** *$\mathcal{P}$* **MAT** *$\mathcal{A}$ such that (m, y) is the $\mathcal{P}$-chosen point in F.*
(3) *There are admissible preferences for which the chosen point is arbitrarily close to S_F, and other admissible preferences for which the chosen point is arbitrarily close to N_F.*

Propositions 1 and 2 deal with a fixed opportunity set. Often we need predictions of how an agent with given preferences will choose in a new opportunity set. Neoclassical preference theory offers a prediction that follows from GARP (or from convexity and positive monotonicity) in the case of standard budget sets. We will sometimes get weaker predictions and sometimes stronger predictions because we deal with more general opportunity sets and with preferences that are convex but not necessarily monotone in other's income y. The following example illustrates this.

Example 5.1. Figure 4 shows standard budget sets F with $I = 1, p = 1$ (solid line) and G with $I = 2, p = 4$ (dashed line). Suppose that a player with preferences $\mathcal{P}$ picks (m_F, y_F) from F. What can we predict about his choice (m_G, y_G) from G? If it happens that (m_F, y_F) is not in G then neoclassical preference theory tells us nothing about (m_G, y_G). Given the increasing benevolence property IB we can make a prediction: (m_G, y_G) lies on the sub-segment southeast of the point (m, y_F) on the G budget line, i.e., $y_G \leq y_F$. This is a consequence of part 2a of the next Proposition.

The result in Example 5.1 can often be strengthened in nonlinear opportunity sets. The point chosen in one opportunity set can be compared to points east of it in another opportunity set using IB, as in part 2b of the next Proposition. As shown in part 3 of the next Proposition, using IB together with $w_y \leq 0$, we can obtain even tighter bounds on choice by constructing a point Z which solves $\text{NTP}_{\partial F}(X) = \text{NTP}_{\partial G}(Z)$. (The Appendix shows how to extend the definition so that Z is well defined even with corners and kinks at which NTP is not single valued.) We say that $Z = (m_Z, y_Z)$ is southeast of $X = (m_X, y_X)$ if and only if $y_Z \leq y_X$ and $m_Z \geq m_X$, and Z is northwest of X if both inequalities are reversed.

Figure 5 illustrates the construction of Z and the main implications of the next Proposition. Part 1 of the Proposition is simply standard revealed preference. Part 2 uses IB to compare WTP at points directly east or west of each other, while part 3 compares points with the same WTP in different opportunity sets.

Proposition 3. *Let a player with strictly convex preferences $\mathcal{A}$ choose $X = (m_F, y_F)$ from opportunity set F and choose $W = (m_G, y_G)$ from opportunity set G. Then:*

(1) *if $X \in G$ then $W \in G \setminus F$ or $W = X$.*
(2) *Let $Y = (\widehat{m}, y_F) \in \partial_E G$ have maximal $\widehat{m}$, and suppose preferences $\mathcal{A}$ satisfy IB. Then*
   (a) *$y_G \leq y_F$ if $NTP_{\partial F}(X) \leq NTP_{\partial G}(Y)$ and $\widehat{m} \leq m_F$, and*
   (b) *$y_G \geq y_F$ if $NTP_{\partial F}(X) \geq NTP_{\partial G}(Y)$ and $\widehat{m} \geq m_F$.*
(3) *Let $Z = (m_Z, y_Z) \in \partial_E G$ solve $NTP_{\partial F}(X) = NTP_{\partial G}(Z)$, and suppose preferences $\mathcal{A}$ satisfy IB and $w_y \leq 0$. Then*
   (a) *$y_G \geq y_Z$ if Z is southeast of X,*
   (b) *$y_G \leq y_Z$ if Z is northwest of X.*

Propositions 1-3 do not invoke Axioms R and S. We now shall see that Axiom R effects can either reinforce or offset the standard revealed preference predictions, depending on the first mover's generosity. The next example also highlights unique predictions arising from Axiom S.

Example 5.2. Suppose that there is a first mover (FM) who picks one of the two standard budget sets for the second mover (SM) in the previous example. Since G **MGT** F, Axiom R implies that the SM's choice $W \in G$ lies northwest of the point (m_G, y_G) predicted by convexity of preferences and the IB property; since (m_G, y_G) is predicted to be southeast of (m, y_F) our model has no testable implication in this instance. Recall that neoclassical preference theory also has no testable implication when (m_F, y_F) does not belong to G. If the FM instead chooses F then Axiom R implies that the choice X lies southeast of (m_F, y_F) whereas neoclassical preference theory predicts that $X = (m_F, y_F)$. Axiom S implies that the choice X^* when the status quo is F lies southeast of the choice X^o when the FM has no choice, and that the choice X^c when the status quo is G lies even further southeast. In contrast, neoclassical preference theory assumes preferences are fixed and therefore predicts $X^c = X^* = X^o$.

6. Diagnostic Tests of Axioms R and S with Investment and Dictator Game Data

Building on Example 5.2, one could design an experiment to test the theory using two player sequential move games involving standard budget sets that are ordered by **MGT**. We will, instead, use existing data from experiments with the Investment and Dictator games. (In the Dictator game, the experimenter gives the SM her opportunity set; the FM has no say in the matter.) These games are better suited to testing behavioral implications of Axioms R and S, as summarized in the following Proposition.

Proposition 4. *Let the FM in the Investment game choose F_s as the SM's opportunity set, and let $r(s)$ be the SM's response. Also let the SM be given the same opportunity set F_s in a Dictator game, and let $r^o(s)$ be his response there.*

(1) *If SM's preferences $\mathcal{A}$ are fixed and satisfy IB, then $r^o(s)$ increases in s.*
(2) *If SM's preferences satisfy Axiom R and IB, then $r(s)$ increases more rapidly in s than does $r^o(s)$.*
(3) *If SM's preferences also satisfy Axiom S, then $r(s) \geq r^o(s)$ for all feasible s.*

Proposition 4 leads to a diagnostic test of Axioms R and S. Our model would be falsified by observations if, contrary to parts 1 and 2, SMs return more in either game when they get s than when they get $s' > s$; or if, contrary to part 3, SMs return more in a Dictator game than in an Investment game with the same opportunity sets F_s.

Using a double-blind protocol, Cox (2004) gathered data from a one-shot Investment game (Treatment A) with 32 pairs of FMs and SMs. Cox also reported parallel data from a Dictator game (Treatment C) with another 32 subject pairs in which the dictators ("SMs") were given exactly the same opportunity sets by the experimenter as were given to SMs by the FMs in the Investment game. In both treatments, the choices s and r were restricted to integer values but the conclusions of Proposition 4 still hold.

To test the predictions, construct the dummy variable $D = 1$ for Treatment C. Regress the SM choice r on the amount sent s and its interaction with D, using a censored regression to account for the limited range of SM choices (r

$\in [0, 3s]$).[3] The estimated coefficient for s is 0.58 ($\pm$ standard error of 0.22) with one-sided p-value of 0.006, consistent with reciprocity and parts 1 and 2 of Proposition 4. The estimated coefficient for $D \times s$ is -0.69 (± 0.32, $p = 0.018$), consistent with Axiom S and part 3 of Proposition 4. Since the coefficient sum is statistically indistinguishable from 0, the convexity prediction in part 1 of Proposition 4 is neither supported nor contradicted.

The above estimation uses observations for all amounts sent s. We here confirm the Axiom S tests result by direct hypothesis tests using a subset of the data with sufficient observations for paired tests: $s = 5$ (with 7 observations in each treatment) and $s = 10$ (with 13 observations in each treatment). The Mann-Whitney and t-test both reject the null hypothesis of no difference between the amounts returned in favor of the strict Axiom S alternative hypothesis that returns are larger in Treatment A. The one-sided p-values for the t-test (respectively the Mann-Whitney test) are 0.027(0.058) for the $s = 5$ data and are 0.04(0.10) for the $s = 10$ data.[4]

7. Tests with Carrot and Stick Game Data

Carrot and Stick games support within-game direct tests of our model and suggest one across-games test. The following proposition draws out the implications of these games.

Proposition 5. *Let the FM in the Stick, Carrot or Carrot-Stick game choose F_s as the SM's opportunity set, and let $r(s)$ be the SM's response.*

(1) *If SM's preferences $\mathcal{A}$ are fixed and satisfy IB, then $r(s)$ increases in s.*
(2) *If SM's preferences satisfy Axiom R and IB, then in the Stick game $r(s)$ increases more rapidly in s than for fixed preferences.*

The model would be falsified by data for any of these games in which SMs chose larger returns $r(s)$ for smaller amounts s sent by the FM. The model suggests that for a given s, smaller (or more negative) returns r should be observed in the Stick game than in Carrot-Stick game. The reasons are twofold. First, comparing the opportunity set F_s^S for given s in Stick to that in Carrot-Stick (F_s^{CS}), one sees that F_s^S **MGT** F_s^{CS}. The **MGT** ordering across games suggests that reciprocity will boost r in the Stick game above its

[3]The constant is set equal to zero because this is implied by the experimental design restriction that SMs cannot return more than they receive from FMs.
[4]Figure 3 in Cox (2004) showing data from Treatments A and C contains a couple of errors. A file with (correct) data from the two treatments is available upon request to the author.

value in the Carrot-Stick game. Second, comparing parts 1 and 2 of the last proposition, one sees that reciprocity boosts r in the Stick game but not in the other two.

Andreoni et al. (2003) report data from Carrot, Stick and Carrot-Stick games, each with 30 pairs of FMs and SMs randomly matched over 10 periods. They focus on choices in the last 5 periods and so shall we.[5] The SM's opportunity set has a kink at $r = 0$ in all three games; 67%, 57%, and 41% of the SM choices are at the kink, respectively, in the Stick game, Carrot game, and Carrot-Stick game. But the kink has different implications across games because FM choices differ across games. Figure 6 shows the percentages of constrained ($r = 0$) responses in the three games for two focal FM choices of the minimum allowable amount sent ($s = 40$) and the equal-split amount sent ($s = 120$).

We want to compare SM choices r across games holding the FM choice s constant, and also want to estimate the impact of s on r in each game. The kinks and resulting returns of zero lead us to separate the data into two parts corresponding to the data presentation in Figures 5 and 6 in Andreoni, et al. (2003): a Stick Regime with choices $r \leq 0$, and a Carrot Regime with choices $r \geq 0$. The Carrot-Stick data are included in both regimes and are indicated by the dummy variable DCS. We use 2-sided tobit estimators since the lower bound in Stick Regime also binds occasionally, as does the upper bound in the Carrot Regime. Random individual subject effects help control for heterogeneous preferences across subjects.

Table I reports the results. Consistent with the predictions from Proposition 5, the amount sent s ("send") has a significantly positive impact in all games and regimes. The estimate 0.36 in the Stick Regime indicates that on average a FM who sends 100 more in Stick will increase r and thus increase his gross payoff by $5 \times 36 = 180$, for a net gain of 80. In Carrot-Stick the estimated marginal impact in this regime is $0.36 + 0.26 = 0.62$, significantly larger at the 3% level, but the intercept is significantly more negative, at $-31.9 - 23.3 = -55.2$. The estimated return function is $r^S(s) = -31.9 + 0.36s$ for Stick, which lies everywhere above its Carrot-Stick counterpart $r^{CSS}(s) = -55.2 + 0.62s$ in the $r \leq 0$ regime. Thus the estimates are consistent with the model's informal across-games implication.

The table reports similar results for the Carrot regime. Again as predicted, the amount sent s by the FM has a significantly positive marginal impact on

[5]Spot checks indicate no substantial changes in results when all 10 periods are included.

the amount returned by the SM. The 0.41 coefficient in the Carrot game is not distinguishable from that in the Carrot-Stick game in the same regime, nor from its Stick counterpart. The model offers no hint about the relative positions of the return functions in this regime, but the data show that the Carrot function $r^C(s)$ is significantly higher than the Carrot-Stick function $r^{CSC}(s)$ in the $r \geq 0$ regime.

TABLE I

TOBIT PANEL REGRESSIONS WITH RANDOM EFFECTS FOR DEPENDENT VARIABLE r

	Stick Regime[a]	Carrot Regime[a]
constant	-31.91	-58.10
	(0.00)	(0.00)
DCS	-23.34	-48.25
	(0.03)	(0.01)
send	0.36	0.41
	(0.00)	(0.00)
DCS×send	0.26	-0.09
	(0.03)	(0.23)
(left,uncensored,right)[b]	(15,67,218)	(179,112,9)

[a]Data are from the last 5 periods of Carrot and/or Stick games (Andreoni et al, 2003). One-sided p-values are shown in parentheses.

[b]The last row shows the number of (left censored obervations, uncensored observations, right censored observations).

8. TESTS WITH STACKELBERG DUOPOLY DATA

A limitation of the preceding applications is that data come from games with opportunity sets with linear Eastern boundaries, so SMs face a constant NTP. The standard Stackelberg game in Example 3.4 escapes this straightjacket. Recall that smaller output choices by the Stackelberg Leader create **MGT** opportunity sets for the Follower. Axiom R says that this will induce **MAT** preferences in the Follower. Due to the higher WTP, it seems that the Follower should choose points on the Eastern boundary with higher NTP, hence larger y, by reducing output.

It's not quite that simple, however. We must also take into account preference convexity, and also the changing curvature of the opportunity set. The

next proposition sorts out these effects and expresses them in terms of the Follower's deviation from selfish best reply (the prediction of standard duopoly theory).

Proposition 6. *In the Stackelberg game of Example 3.4 let* $Q_D(q_L) = q_F - q_F^o$ *be the deviation of the Follower's output choice* q_F *from the selfish best reply* $q_F^o = 12 - \frac{1}{2}q_L$ *when the Leader chooses output* q_L*. One has*

$$\frac{dQ_D}{dq_L} = -\frac{1}{2}w - \frac{dw}{dq_L}q_L$$

where $w = WTP(Mq_F, Mq_L)$ *is willingness to pay at the chosen point. Furthermore,*

(1) *If Follower's preferences* $\mathcal{A}$ *are fixed and linear, then* w *is constant with respect to* q_L *and* $\frac{dQ_D}{dq_L}$ *is positive if and only if preferences at the chosen point are malevolent.*
(2) *If Follower's preferences* $\mathcal{A}$ *are fixed, satisfy IB and* $w \leq 1$*, then* w *is decreasing in* q_L *and* $\frac{dQ_D}{dq_L}$ *contains an additional positive term.*
(3) *If Follower's preferences satisfy Axiom R strictly, then* w *is decreasing in* q_L *and* $\frac{dQ_D^r}{dq_L}$ *contains an additional positive term.*
(4) *If Follower's preferences satisfy Axiom S strictly, then* w *is decreasing in* q_L *and* $\frac{dQ_D^s}{dq_L}$ *has an additional positive (negative) term if the status quo is smaller (larger) than* q_L*.*

Proposition 6 shows that an increase in q_L has three different effects:

- A reciprocity effect, items (3) - (4) in the Proposition. If Axiom R holds strictly, then the less generous opportunity set decreases the Follower's WTP, increasing q_F and $q_D = Q_D(q_L)$. Axiom S moderates or intensifies this effect, depending on the status quo.

- A preference convexity (or substitution) effect, item (2) in the Proposition. The choice point is pushed west, where WTP is less, again increasing q_D.

- An opportunity set shape effect (in some ways analagous to an income effect), item (1) in the Proposition. The curvature of the parabola decreases. Holding w =WTP constant, q_D increases when the Follower is malevolent ($w < 0$, hence $q_D > 0$), and decreases when the Follower is benevolent ($w > 0$, hence $q_D < 0$).

A parametric example may clarify the logic. For given $q_L \in [0, 24]$, the Follower's choice set is the parabola $\{(m, y) : m = Mq_F, y = Mq_L, M = 24 - q_L - q_F, q_F \in [0, 24 - q_L]\}$, with NTP$= -\frac{dm/dq_F}{dy/dq_F} = \frac{24 - q_L - 2q_F}{q_L}$. Suppose that the Follower has fixed Cobb-Douglas preferences represented by $u(m, y) = my^\theta$,

so WTP is $\theta m/y = \theta q_F/q_L$. Solving NTP=WTP, one obtains $q_F = Q(q_L|\theta) = (24 - q_L)/(2 + \theta)$. Noting that the selfish best reply is $q_F^o = Q(q_L|0)$, one obtains a closed form expression for the deviation, $q_D = -\frac{\theta}{4+2\theta}(24 - q_L)$. For fixed θ positive (benevolent preferences) or smaller than -2 (pathologically malevolent preferences), the deviation is negative but increasing in the Leader's output; the opposite is true when θ is negative but larger than -2 (moderately malevolent). This is the combined impact of the convexity (or substitution) and shape (or income) effects noted above. Of course, reciprocity effects will decrease θ and hence increase q_D.

We test predictions obtained from Proposition 6 on the Stackelberg duopoly data of Huck, Müller, and Normann (2001, henceforth HMN). The parameters are exactly as in Example 3.4 with integer output choices. The data consist of 220 output pairs (q_L, q_F) by 22 FMs (or Leaders) choosing $q_L \in \{3, 4, 5, \ldots, 15\}$ randomly rematched for 10 periods each with 22 SMs (or Followers) who choose $q_F \in \{3, 4, 5, \ldots, 15\}$. The WTP can be inferred at a chosen point (q_L, q_F) by the NTP at that point, $(24 - 2q_F - q_L)/q_L$.

Table II reports the test results. All observations reveal $w \leq 1$, as assumed in Proposition 6. To check for asymmetric responses to large and small FM choices (relative to the Cournot choice $q_L = 8$), we define the dummy variable $DP = 1$ if $q_L \leq 8$. All columns in the table report panel regressions with individual subject fixed effects. The first column, with dependent variable WTP$\times$100, firmly rejects the hypothesis of benevolent linear and fixed preferences: the coefficient for q_L is significantly negative, not positive. In view of part 1 of the Proposition, the second column, with dependent variable Q_D, confirms this result. We infer that Q_D is an increasing function of FM output q_L, consistent with convexity and reciprocity, in view of parts 2 and 3 of the Proposition. The last column reports that there is a stronger response to "greedy" FM choices in excess of the Cournot output 8 than to "generous" FM choices below or equal to output 8.

TABLE II
PANEL REGRESSIONS WITH FIXED EFFECTS[a]

Dep.Variable	WTP×100	q_D	q_D
q_L	-4.57	0.32	0.23
	(0.00)	(0.00)	(0.001)
DP×q_L			-0.11
			(0.017)
constant	21.56	-1.88	-0.70
	(0.012)	(0.000)	(0.177)

[a]Data consist of 220 choices by 22 Followers in HMN experiment. One-sided p-values are shown in parentheses.

9. Diagnostic Tests of Reciprocity with Stackelberg Mini-Game Data

The Stackelberg duopoly game data do not permit tests of some of our most distinctive predictions. All FMs (Leaders) have the same choice set, eliminating variability that could help separate the convexity effect from the reciprocity effect. Also, due in part to differing experiences, SMs may have different views on the generosity of a given output choice q_L. In order to overcome these limitations while preserving the nice parabolic shape of the SM choice sets, we introduce a new Stackelberg variant.

Example 9.1. Stackelberg Mini-Game. Take the otherwise standard Stackelberg duopoly game in Example 3.4, but restrict the Leader (FM) to a binary output choice, $q_{_L} \in \{x, z\}$, where $0 < x < z < 24$.

The idea here is to manipulate the Leader's choice set in order to obtain a direct test of reciprocity. In one situation, a given output choice can be the smaller one allowed (hence the most generous to the Follower) and in another situation the same choice can be the larger one (hence the least generous). If a given Follower reacts differently in the two situations, it must be due to reciprocity effects, since by holding the Leader's output constant we have eliminated convexity and shape effects. Formally,

Proposition 7. *In the Stackelberg Mini-Game of Example 9.1, suppose the Leader has restricted output choices $q_{_L} \in \{x, s\}$ in situation (a) and $q_{_L} \in \{s, z\}$ in another situation (b), where s is strictly between x and z. Suppose the*

Leader chooses s in both situations and the Follower choses $Q_D^a(s)$ in situation (a) and $Q_D^b(s)$ in situation (b). If the Follower's preferences satisfy Axioms R and S, then $Q_D^a(s) \geq Q_D^b(s)$, and at each possible Follower choice q_F, $WTP^a(Mq_F, Ms) \leq WTP^b(Mq_F, Ms)$.

Thus, contrary to standard revealed preference theory, the model predicts that the Follower's choice in a fixed opportunity set F depends in a specific way on the alternatives *not* chosen by the Leader. Our model would be falsified by observations if Followers choose larger quantities or reveal higher WTPs when Leaders forgo $z > s$ to choose s than when Leaders forgo $x < z$ to choose s.

In our new Stackelberg mini-game experiment, each subject in the FM role twice chooses $q_L \in \{6, 9\}$ and twice chooses $q_L \in \{9, 12\}$ without feedback. Each subject in the SM role is then paired simultaneously with four different FMs and chooses an integer value of $q_F \in \{5, 6, ..., 11\}$ with no feedback. The corresponding payoffs (m, y) are clearly displayed. Subjects are paid for one of the four choices, selected randomly at the end of the session. The "double blind" procedures are detailed in the instructions to subjects, reproduced in Appendix B.

To infer how individual subjects respond to reciprocity concerns, we turn again to panel regressions with individual subject fixed effects. The second column in Table III reports that, consistent with Proposition 7, SMs' average WTP decreased by almost 8 cents per dollar when $q_L = 9$ was the less generous choice (indicated by $D9 = 1$). The second column reports the same data in a different way: the output deviation increased by 0.34 on average, significant at the $p = 0.016$ level (one-sided) when $D9 = 1$. Since the opportunity set F_9 is constant in these 72 data points, the result cannot be due to convexity or shape effects; it must be pure reciprocity. The last column of Table III reports regressions for q_D for the entire data set, using the additional dummy variable $D12$, which takes value 1 if $q_L = 12$, and 0 otherwise.[6] The signs of all coefficient estimates are consistent with Axioms R and S and convexity.

[6]We omit here a dummy variable that takes value 1 for $q_L = 6$ because there are only five such observations. When the dummy is included, the coefficient estimate has the predicted sign but of course is insignificant statistically, while the other coefficient estimates change only slightly.

TABLE III
PANEL REGRESSIONS WITH FIXED EFFECTS FOR
STACKELBERG MINI GAME DATA[a]

	$w\times100$ (q_L= 9)	q_D (q_L= 9)	q_D
D9	-7.65	0.34	0.32
	(0.008)	(0.008)	(0.013)
D12			0.37
			(0.028)
constant	-5.93	0.27	0.19
	(0.007)	(0.007)	(0.046)
Nobs (gr)[b]	72(24)	72(24)	91(24)

[a]One-sided p-values are shown in brackets.

[b]Nobs is the total number of observations (gr is the number of groups).

10. DISCUSSION

Neoclassical theory (e.g., Hicks (1939), Samuelson (1947)) clarified and unified earlier work on how opportunities and preferences jointly determine outcomes for *homo economicus*. The present paper applies those now-classic ideas to social preferences. We focus on need to pay (NTP), the reciprocal of the marginal rate of transformation of own income into others' income, and willingness to pay (WTP), the reciprocal of the marginal rate of substitution between own income and others' income. Increasing WTP along indifference curves is simply convexity, and convex altruistic preferences provide a unified account of several social motives previously considered separately, such as efficiency, maximin, and inequality aversion.

We develop a theory of reciprocal altruism: how choices by one player shift preferences of another player and determine outcomes for *homo reciprocans*. We say that one opportunity set G is more generous to person X than another opportunity set F, and write G **MGT** F, if the maximum income in G for person X exceeds his maximum in F, and does so by more that the corresponding income difference for the other player. We say that one set of preferences is more altruistic than (**MAT**) another if it has a larger WTP at every point. We formalize reciprocity as a **MAT**-tilt in preferences following a **MGT** choice by others. The definitions apply to malevolent (WTP < 0) as well as benevolent (WTP > 0) preferences, and automatically combine positive and negative reciprocity.

Convexity and reciprocity are quite different formally and conceptually, but we show that empirical work has a natural tendency to confound the two notions. The problem is simply that more generous behavior by a first mover tends to push the second mover's opportunities southeast, towards larger income for the second mover and smaller income for the first mover. Convexity typically implies greater WTP as one pushes southeast, even when there is no **MAT**-shift in preferences due to reciprocity.

Axiom R and Axiom S set revealed altruism theory apart from neoclassical preference theory. In neoclassical theory, my preferences are an individual characteristic that is independent of your prior actions that help or harm me. In contrast, Axiom R asserts that more generous choices by you induce more altruistic preferences in me. Axiom S further asserts that my induced preferences are more altruistic when your generous choice is an act of commission (upsetting the status-quo) than when it is an act of omission (upholding the status-quo), and that this reciprocity effect is even weaker when you are unable to alter the status quo. The theory incorporates negatively-reciprocal altruism: less generous choices by you induce less altruistic preferences in me, where "less altruistic" can mean "more malevolent."

Several theoretical propositions develop the observable consequences of neoclassical properties such as convexity and the new reciprocity Axioms. We show that more northerly choices on the Eastern boundary of an opportunity set reveal more altruistic (or less malevolent) preferences. For fixed preferences, choices in one opportunity set reveal bounds on preferences that translate into bounds on choices in other opportunity sets. For reciprocal preferences, a first mover's choice of a more or less generous opportunity set translates into bounds on a second mover's choice, and the bounds are contingent on the status quo ante. We derive propositions tailored to a set of well-known two player games: Investment, Dictator, Carrot and/or Stick, and Stackelberg duopoly. The tailored propositions sort out the separate effects of the neoclassical properties and the new Axioms. The paired Investment and Dictator games provide a diagnostic test of the implications of both Axiom R and Axiom S. The new Stackelberg mini-game provides a diagnostic separation of the implications of convexity and reciprocity.

Finally, to illustrate the empirical content of the theory, we examine three existing data sets and one new data set. Existing data from Investment and Dictator experiments reject null hypotheses inconsistent with Axioms R and

S in favor of alternative hypotheses consistent with the Axioms (and convexity). Existing data from the Stick game and the Carrot and Stick game support implications of Axiom R (and convexity). Existing data from a Stackelberg duopoly experiment confirm reciprocity/convexity effects and suggest a stronger negative response to greedy behavior than the positive response to generous behavior. Data from a new experiment with the Stackelberg mini-game confirm that reciprocity has a significant impact even when convexity effects are held constant. The Stackelberg mini-game brings out a novel feature of the new theory: contrary to standard revealed preference theory, revealed altruism theory explains how alternatives *not* chosen by another can affect one's own choice.

Theoretical clarification sets the stage for further empirical work. One can now refine earlier empirical studies that examine the combined effects of altruism and reciprocity. Such work should shed light not only on the extent to which typical human preferences depart from selfishness but also on the extent to which such departures are altered by experiencing generous or ungenerous behavior.

Further theoretical work is also in order. We consider two versions of the "more generous than" relation but yet other versions may have implications that are stronger (or just different). For example, generosity might be defined in terms of players' utilities rather than in terms of material payoffs (although this would compromise observability). Other open theoretical questions concern Axiom S, which invokes the status quo to distinguish between acts of commission and omission, and between generous and greedy acts. But what does it take for a particular act to become generally recognized as the status quo? What if an act has beneficial short run impact but is harmful in the long run? Answers to these and other questions await further theoretical development.

Department of Economics and Experimental Economics Center (ExCEN), Andrew Young School of Policy Studies, Georgia State University, P.O. Box 3992, Atlanta, GA 30302-3992, U.S.A.; jccox@gsu.edu,

Department of Economics, 417 Building E2, University of California, Santa Cruz, CA 95064, U.S.A.; dan@ucsc.edu,

and

Department of Economics and Experimental Economics Center (ExCEN), Andrew Young School of Policy Studies, Georgia State University, P.O. Box 3992, Atlanta, GA 30302-3992, U.S.A.; vsadiraj@gsu.edu.

Appendix A. Mathematical proofs and derivations

A.1. ***Properties of Preferences.*** Recall that preferences over bundles $(m, y) \in \Re^2_+$ are admissible if they can be represented by a twice continuously differentiable (smooth) utility function u such that $\partial u(m,y)/\partial m = u_m > 0$, $\forall (m, y) \in \Re^2_+$ (m-monotone) and the set $\{(m, y) \in \Re^2_+ : u(m, y) \geq c\}$ is convex for all $c \in \Re$ (convex). Recall also that willingness to pay is $w = w(m, y) = \text{WTP}(m, y) = u_y/u_m$.

It will be helpful to express convexity in terms of the curvature of indifference curves. At a given point, curvature has absolute value $|K| = 1/R$, where R is the radius of the circle that is second-order tangent to the curve at the given point. Let θ denote the angle of the tangent to the indifference curve with the negative y-direction. The signed curvature is $K = d\theta/ds$ where $s(t) = \int_0^t \sqrt{m'^2(x) + y'^2(x)}dx$ is arclength along the indifference curve (e.g., Protter and Morrey, (1963, p. 394)).

Preferences are positively monotonic in m; hence upper contour sets are on the right of indifference curves in (m, y) space. The convexity of upper contour sets implies that w decreases as we move up along the indifference curve. The first lemma verifies this intuition and obtains other useful characterizations.

Lemma A.1. *The following properties are equivalent for smooth m-monotone preferences on $\Re^2_+$:*
(a) They are convex.
(b) Their indifference curves everywhere have negative (or zero) curvature.
(c) $w_m w - w_y \geq 0$.

Proof: Note that along the indifference curve $\theta = \arctan(dm/dy) = \arctan(-w)$. Into the definition $K = d\theta/ds$, insert $d\theta = -d(w)/(1 + w^2)$, $ds = \sqrt{dm^2 + dy^2}$ and (holding u constant) $-dm/dy = w$ to get

$$K = \frac{1}{\sqrt{w^2 + 1}^3} \frac{dw}{dy}. \tag{A.1}$$

Since the expression inside the radical is positive, the sign of K is that of $\frac{dw}{dy}$. The upper contour set at a point (m_o, y_o) with $u(m_o, y_o) = c$ lies on the right or on the tangent hyperplane if and only if $(dw/dy)|_{u(m,y)=c} \leq 0$, as can be seen, e.g., from a straightforward adaptation of Protter and Morrey (1963).

Hence conditions (a) and (b) are equivalent. To verify the equivalence of (b) and (c), simply substitute $dw/dy = w_m dm/dy + w_y$ and $dm/dy = -w$ into (A.1) to obtain

$$K = -\frac{ww_m - w_y}{\sqrt{w^2+1}^3}. \tag{A.2}$$

Q.E.D.

Lemma A.2. *$(d(NTP)/dy)|_{(m,y)\in\partial F} \geq 0$ at every regular boundary point of an opportunity set.*

PROOF: The reasoning is the same as in the previous Lemma. Along the boundary

$$K = \frac{1}{\sqrt{NTP^2+1}^3}\frac{d(NTP)}{dy}. \tag{A.3}$$

Thus $K = d\theta/ds$ has the same sign as $d(NTP)/dy$. Our feasible opportunity set F lies on the left or on the tangent hyperplane at a point from the boundary ∂F. Hence, as y increases the boundary is turning left, so θ increases and (by A.3) NTP increases.

Q.E.D.

The next Lemma characterizes homotheticity in order to facilitate comparisons to the weaker properties used in the Propositions.

Lemma A.3. *The following are equivalent:*
(a) Preferences are homothetic on $\Re^2_+$.
(b) $w =$WTP is constant along every ray $R_r = \{(t, tr) : t > 0\} \subset \Re^2_+$.
(c) $w_m + w_y r = 0$ along every ray R_r, $r > 0$.

PROOF: By definition, preferences are homothetic if and only if they can be represented by a utility function $u(m, y)$ whose ratio of partial derivatives u_m/u_y depends only on the ratio m/y (e.g., Simon and Blume (1994, p. 503)). Thus condition (a) implies that $w = u_y/u_m$ is constant along the ray with $r = m/y$ and so condition (b) must hold. In turn, condition (b) implies that along that ray $0 = dw/dt = w_m dm/dt + w_y dy/dt = w_m + w_y r$, establishing condition (c). Since rays with $r > 0$ foliate $\Re^2_+ \setminus (0, 0)$, condition (c) implies that w and hence u_m/u_y depend only on $r = m/y$, i.e., (a) must hold.

Q.E.D.

Definition 3. Preferences are rather malevolent (resp. not very malevolent) on a domain D if $w \leq w_y/w_m$ (resp. $w \geq w_y/w_m$) holds at all points in $D \subset \Re^2_+$.

Lemma A.4. *(a) If admissible preferences are homothetic on $\Re^2_+$ then they are IB.*
(b) Convexity on $\Re^2_+$ is equivalent to IB for preferences that are not very malevolent, and is equivalent to $w_m \leq 0$ for preferences that are rather malevolent.

PROOF: For part a we need to show that $w_m(m, y)$ is non-negative. It sufficies to show that the sign of $w(m + \delta, y) - w(m, y)$ is the same as the sign of δ, for all δ. If $\delta > 0$ then $(m + \delta, y)$ is on a ray $(R_{y/(m+\delta)})$ with a smaller slope than the ray through (m, y). This, convexity and homotheticity imply that $w(m+\delta, y) \geq w(m, y)$. Similarly, $w(m+\delta, y) \leq w(m, y)$ for negative δ. For part b, recall from Lemma A.1 that convexity is equivalent to $w_m w - w_y \geq 0$. But this is equivalent to $w_m \geq 0$ ($w_m \leq 0$) if $w \geq w_y/w_m$ ($w \leq w_y/w_m$).

Q.E.D.

To see the bite of the assumptions, consider preferences represented by $u(m, y) = m^r/r - y^{(1+r)}/(1 + r)$. For $r > 1$ these preferences are IB but neither convex nor homothetic. For $r \in (0, 1)$, however, they are convex but neither IB nor homothetic.

A.2. ***Proof that MAT is a Partial Ordering.*** The properties of reflexivity and transitivity are inherited from the reflexivity and transitivity of the real ordering $\geq$. The antisymmetry property follows from Hicks' Lemma (Hicks (1939, Appendix)): if preferences have the same MRS (or WTP) everywhere in a domain D then they are the same on that domain.

Q.E.D.

A.3. ***Proof that MGT is a Partial Ordering.*** Reflexivity, antisymmetry and transitivity all are inherited from the corresponding properties of the real ordering $\geq$.

Q.E.D.

A.4. ***Proof that Stackelberg Follower's opportunity sets are MGT-ranked.*** The Follower's opportunity set F_{q_L} has Eastern boundary $\{(m,y) : m = Mq_F, y = Mq_L, q_F \in [0, T-q_L]\}$ where $M = T - q_L - q_F$. Along this boundary NTP is given by

$$NTP = -\frac{dm/dq_F}{dy/dq_F} = \frac{T - q_L - 2q_F}{q_L}.$$

Note that NTP varies smoothly from positive to negative values as increasing q_F passes through $q_F^o = T/2 - q_L/2$, the selfish best response. To see that a smaller output by the Leader produces a MGT opportunity set for the Follower, first note that $y^*_{F(q_L)} = (T - q_L)q_L$ is obtained when $q_F = 0$ and that $m^*_{F(q_L)} = \frac{1}{4}(T - q_L)^2$ is obtained from the standard (selfish) reaction function q_F^o. To verify condition (a) in the **MGT** definition, let $q_L^{'} \in (q_L, T - q_L)$ and note that $m^*_{F(q_L)} - m^*_{F(q_L^{'})} = \frac{1}{4}(2T - q_L - q_L^{'})(q_L^{'} - q_L) > 0$. Condition (b) follows from $y^*_{F(q_L)} - y^*_{F(q_L^{'})} = (q_L + q_L^{'} - T)(q_L^{'} - q_L) \leq 0$.

Q.E.D.

A.5. ***Examples of MGT-ordered Opportunity Sets.***

Example A.5. Ring test (Liebrand (1984); see also (Sonnemans, van Dijk and van Winden (2005)). Let $F(R) = \left\{(m,y) \in \Re^2_+ : m^2 + y^2 \leq R^2\right\}$ for given $R > 0$. On the circular part of the boundary, NTP is y/m and the curvature is $1/R$. Straightforwardly, $F(R)$ **MGT** $F(R^{'})$ if $R > R^{'}$.

Example A.6. Ultimatum game (Güth, Schmittberger, and Schwarze (1982)). The responder's opportunities in the \$10 ultimatum game consist of the origin $(0,0)$ and (due to our free disposal assumption) the horizontal line segment from $(0, 10-x)$ to $(x, 10-x)$. This set is not convex so it doesn't qualify as an opportunity set by our definition. Its convex hull, however, is the opportunity set in the Convex Ultimatum game (Andreoni, Castillo and Petrie (2003), which is identical to that of the Power to Take game in the following example.

Example A.7. Power to Take game (Bosman and van Winden (2002)). The "take authority" player chooses a take rate $b \in [0,1]$. Then the responder with income I chooses a destruction rate $1 - \delta$. The resulting payoffs are $m = (1-b)\delta I$ for the responder and $y = b\delta I$ for the take authority. Thus, with free disposal the responder's opportunity set is the convex hull of three points $(m,y) = (0,0), (0,bI)$ and $((1-b)I, bI)$. Along the Eastern boundary NTP is constant at $(b-1)/b$ and the curvature is 0. To verify the strict **MGT**

ranking, let $b' > b \geq 0$ produce SM opportunity sets F and G respectively, so $m_F^* = (1-b')I$ and $y_F^* = b'I$. Then $m_G^* - m_F^* = (b'-b)I > 0 > (b-b')I = y_G^* - y_F^*$. The first inequality confirms part a of the definition and the entire string confirms part b.

Example A.8. Ultimatum mini-games (Gale, Binmore, and Samuelson (1995), Falk, Fehr and Fischbacher (2003)). In the notation of the previous example, the FM in these games chooses between $b = 0.8$ and one other value, either $b = 0.5$ in the 5/5 game, or $b = 0.2$ in the 2/8 game, or $b = 0.8$ in the 8/2 game, or $b = 1.0$ in the 10/0 game. The previous example shows that the (convexified) opportunity sets are **MGT**-ranked by decreasing b. Axiom R suggests that the SM is more likely to choose $(0,0)$ (reject the ultimatum) rather than $((1-b)I, bI)$ (accept) when the FM's choice of b was less generous. Hence rejections of the $b = 0.8$ proposal should be more frequent when the alternative was $b = 0.5$ or $b = 0.2$ rather than $b = 1.0$. Axiom S suggests that the responses would be muted when the alternative was $b = 0.8$ (i.e., no choice). The data are consistent with these predicitions; see Cox, Friedman and Gjerstad (2007) for a detailed structural analysis.

Example A.9. Moonlighting game (Abbink, Irlenbusch, and Renner (2000)). In this variant of the investment game, the FM sends $s \in [-I/2, I]$ to SM, who receives $g(s) = ks$ for positive s and $g(s) = s$ for negative s. Then the second mover transfers $t \in [(-I+s)/k, I+g(s)]$ resulting in non-negative payoffs $m = I + g(s) - |t|$, and $y = I - s + t$ for positive t and $y = I - s + kt$ for negative t. The second mover's opportunity set is the convex hull of the points $(m,y) = (0,0), (I+g(s)-(I-s)/k, 0), (I+g(s), I-s)$, and $(0, 2I+g(s)-|s|)$. The NTP along the boundary of the opportunity set is 1 above and $-1/k$ below the $t = 0$ locus, is 0 along the y axis, and is ∞ along the m-axis. Again, curvature at all regular boundary points is $K = 0$. It is straightforward to verify that larger s produces higher **MGT**-ranking.

Example A.10. Gift exchange labor markets (Fehr, Kirchsteiger, and Riedl (1993)). The employer with initial endowment I offers a wage $W \in [0, I]$ and the worker then chooses an effort level $e \in [0,1]$ with a quadratic cost function $c(e)$. The final payoffs are $m = W - c(e)$ for the worker and $y = I + ke - W$ for the employer, where the productivity parameter $k = 10$ in a typical game. The worker's opportunity set is similar to the second mover's in the investment game, except that the Northeastern boundary is a parabolic arc instead of a straight line of slope -1. Along this Eastern boundary NTP is $2e$ and the

curvature is $-1/5(4e^2+1)^{3/2}$. Also, if the employer offers a wage in excess of his endowment I then the opportunity set includes part of the quadrant $[m > 0 > y]$. It is straightforward but a bit messy to extend the definition of opportunity set to include such possibilities. Again, one can directly verify that larger W produces higher **MGT**-ranking.

Example A.11. Sequential VCM public good game with two players (Varian (1994)). Each player has initial endowment I. FM contributes $c_1 \in [0, I]$ to the public good. SM observes c_1 and then chooses his contribution $c_2 \in [0, I]$. Each unit contributed has a return of $a \in (0.5, 1]$, so the final payoffs are $m = I + ac_1 - (1-a)c_2$ for SM and $y = I + ac_2 - (1-a)c_1$ for FM. SM's opportunity set is the convex hull of the four points $(m, y) = (0, I-(1-a)c_1), (I+ac_1, I-(1-a)c_1), (aI+ac_1, (1+a)I-(1-a)c_1)$ and $(0, (1+a)I-(1-a)c_1)$. Along the Pareto frontier, NTP is constant at $(1-a)/a$. Once again, a larger contribution c_1 creates **MGT** opportunities for the second mover.

A.6. ***Proof of Proposition* 1.** Suppose that $(m_{\mathcal{A}}, y_{\mathcal{A}}) \notin \partial_E F$. Then by definition of $\partial_E F$ there exists $z > m_{\mathcal{A}}$ such that $M = (z, y_{\mathcal{A}}) \in F$. Positive monotonicity in own payoff implies that M is strictly preferred to $(m_{\mathcal{A}}, y_{\mathcal{A}})$, contradicting the hypothesis that $(m_{\mathcal{A}}, y_{\mathcal{A}})$ is the $\mathcal{A}$-preferred point in F.

Q.E.D.

A.7. ***Proof of Proposition 2 (Theoretical Predictions for Fixed Opportunity Sets).*** By Lemma A.2, NTP increases as y increases along $\partial_E F$.

Part 1. Convexity of F and optimality of $(m_{\mathcal{B}}, y_{\mathcal{B}})$ imply that $\partial_E F$ (including the part north of $(m_{\mathcal{B}}, y_{\mathcal{B}})$) lies in the negative closed halfspace for the tangent line, $H_{\mathcal{B}}$ to the $\mathcal{B}-$indifference curve through $(m_{\mathcal{B}}, y_{\mathcal{B}})$. $\mathcal{B}$ **MAT** $\mathcal{A}$ implies that the tangent line, $H_{\mathcal{A}}$ of the $\mathcal{A}-$indifference curve through the same point $(m_{\mathcal{B}}, y_{\mathcal{B}})$ is a clockwise rotation of $H_{\mathcal{B}}$. Hence, the $\partial_E F-$points north of $(m_{\mathcal{B}}, y_{\mathcal{B}})$ are from the negative halfspace of $H_{\mathcal{A}}$ and from convexity of preferences $\mathcal{A}$ their $\mathcal{A}-$indifference curves are at lower levels than $(m_{\mathcal{B}}, y_{\mathcal{B}})$. Therefore $(m_{\mathcal{A}}, y_{\mathcal{A}})$ must be south of $(m_{\mathcal{B}}, y_{\mathcal{B}})$.

Part 2. By hypothesis, $X = (m, y)$ is south of $(m_{\mathcal{B}}, y_{\mathcal{B}})$ and north of $(m_{\mathcal{A}}, y_{\mathcal{A}})$. Let w_a and w_b denote WTP functions for $\mathcal{A}$ and $\mathcal{B}$ preferences; by admissibility the functions are continous. The desired conclusion is trivial if $w_b(m, y) = w_a(m, y)$ and Lemma A.2 rules out $w_b(m, y) < w_a(m, y)$, so suppose $w_b(m, y) > w_a(m, y)$. To construct the desired preferences $\mathcal{P}$, let

$w_P(Y) = kw_b(Y) + (1-k)w_a(Y)$ where

$$k = \frac{NTP(m,y) - w_a(m,y)}{w_b(m,y) - w_a(m,y)}.$$

Since w_P is continous on $\Re^2_+$, classic theorems assure the existence of a utility function whose WTP is $w_P(Y)$ (Hurewicz (1958, p. 7-10); see also Hurwicz and Uzawa (1971)). Let $\mathcal{P}$ the preferences represented by this utility function. Since the hypothesis implies that $0 < k < 1$, we have $\mathcal{B}$ **MAT** $\mathcal{P}$ **MAT** $\mathcal{A}$. By construction, (m, y) is $\mathcal{P}$-chosen since $w_P(m, y) =$NTP(m, y).

Part 3. Linear preferences with w approaching $-\infty$ $(+\infty)$ yield choices arbitrarily close to S_F (N_F).

Q.E.D.

A.8. ***Proof of Proposition 3 (Theoretical Predictions for Different Opportunity Sets).*** Suppose that X is a regular point from $\partial_E F$. Then $x =$NTP(X) is unique. Let the NTP of points from $\partial_E G$ take values between $[\gamma_*, \gamma^*]$. Z is: N_F, if NTP$(X) > \gamma^*$, S_F, if NTP$(X) < \gamma_*$; otherwise Z is the point of $\partial_E G$ with $x \in$NTP(Z). Such a point exists by the Intermediate Value Theorem and is unique because G is convex. If X is not a regular point then NTP(X) takes values from some $[\delta_*, \delta^*]$. Make the arbitary convention that $x = \delta^*$ and proceed as with a regular point.

Part 1. Follows from standard revealed preference theory (e.g., Varian (1992, p. 131-133)).

Part 2. Clearly $\widehat{m} \leq m_F$ and $WTP_m \geq 0$ imply $WTP(Y) \leq WTP(X)$, while $NTP_{\partial F}(X) \leq NTP_{\partial G}(Y)$, optimality of X (so $WTP(X) = NTP_{\partial F}(X)$) and transitivity together imply that $WTP(Y) \leq NTP_{\partial G}(Y)$. By convexity of $\mathcal{A}$ all points from ∂G north of Y are on lower $\mathcal{A}-$indifference curves than Y so they cannot be W. Thus W must be south of Y, and 2a follows. One obtains 2b in just the same way.

Part 3. Suppose Z is southeast of X. Then

$$WTP(Z) \geq WTP(X) = NTP_{\partial F}(X) = NTP_{\partial G}(Z).$$

where the first inequality follows by assumption whereas the equalities follow from optimality of X and by construction of Z. By convexity of $\mathcal{A}$ all points from ∂G south of Z are on lower $\mathcal{A}-$indifference curves than Z so they cannot be W. That is, W must be north of Z. Likewise for the case with Z northwest of X.

Q.E.D.

A.9. ***Proof of Proposition 4 (Investment Game).*** **Part 1.** Let $r(s)$ be the optimal choice of SM when the FM choice is s and let $X_{F_s} = (10 + 3s - r(s), 10 - s + r(s))$. Let $s^{'} > s$. Proposition 3.2.b tells us that $y_{F_{s'}} \geq y_{F_s}$. This implies that $r(s^{'}) \geq s^{'} - s + r(s) > r(s)$.

Part 2. Applying Axiom R in the argument above, we see that $r(s^{'})$ increases more rapidly in s than for fixed preferences.

Part 3. Axiom S has the indicated impact since, as shown in the previous subsection, F_s is **MGT** ordered by s.

Q.E.D.

A.10. ***Proof of Proposition 5 (Carrot, Stick and Carrot-Stick Games).*** Let $r(s)$ be the optimal choice of SM when the FM choice is s and let $X_{F_s} = (s - |r(s)|, 240 - s + 5r(s))$. Let $s^{'} > s$. The amount return $r(s)$ is non-positive in the Stick Game, non-negative in the Carrot game and it can be both in the Carrot-Stick game.

Part 1. Proposition 3.2.b tells us that $y_{F_{s'}} \geq y_{F_s}$. This implies that $r(s^{'}) \geq (s^{'} - s)/5 + r(s) > r(s)$.

Part 2. Applying Axiom R in the argument above, we see that $r(s)$ increases more rapidly in s than for fixed preferences in the Stick game.

Q.E.D.

A.11. ***Proof of Proposition 6 (Stackelberg Duopoly Game).*** The FOC can be written as $w(q_F, q_L) \equiv$WTP(Mq_F, Mq_L) =NTP= $(24 - 2q_F)/q_L - 1$, which can be rewritten as

$$q_F = 12 - \frac{w(q_F, q_L) + 1}{2} q_L. \tag{A.4}$$

Inserting the definition of Q_D from the statement of the proposition, we obtain

$$Q_D = -\frac{w(q_F, q_L)}{2} q_L. \tag{A.5}$$

Part 1. *Linear preferences.* If Follower's preferences are fixed and linear with WTP= w then differentiation of (A.5) with respect to q_L gives

$$\frac{dQ_D}{dq_L} = -\frac{w}{2}.$$

Part 2. *Convex Preferences.* If Follower's preferences are fixed and convex then

$$\frac{dQ_D}{dq_L} = -\frac{w(q_F, q_L)}{2} - \frac{q_L}{2}\frac{dw(q_F, q_L)}{dq_L}$$

The additional (second) term above is positive because, as we now will verify, $dw(q_F, q_L)/dq_L$ is negative. Indeed,

$$\begin{aligned}\frac{dw(q_F, q_L)}{dq_L} &= w_m\frac{dm}{dq_L} + w_y\frac{dy}{dq_L} \\ &= w_m((-1 - \frac{dq_F}{dq_L})q_F + M\frac{dq_F}{dq_L}) + w_y((-1 - \frac{dq_F}{dq_L})q_L + M)\end{aligned}$$

which after substituting $M = 24 - q_L - q_F$, $q_F = 12 - (w(q_F, q_L) + 1)q_L/2$ and $dq_F/dq_L = -(w(q_F, q_L) + 1)/2 - (dw(q_F, q_L)/dq_L)q_L/2$ and solving for $dw(q_F, q_L)/dq_L$ we get

$$\frac{dw(q_F, q_L)}{dq_L} = \frac{B}{A}$$

where

$$A = 2 + [w_m w - w_y]\, q_L^2 > 0$$

by Lemma (A.1), and

$$\begin{aligned}B &= 24(w_y - w_m) + q_L(1 - w)(w_m - w_y + ww_m - w_y) \\ &= (w_y - w_m)(24 - q_L(1 - w)) + q_L(1 - w)(ww_m - w_y)\end{aligned}$$

Note that $q_F \leq 24 - q_L$ and (A.4) imply that the sign of the second factor in the first term is positive; hence B is negative if and only if

$$\frac{w_y - w_m}{ww_m - w_y} \leq \frac{1 - w}{24/q_L - (1 - w)}$$

The current assumptions of $w \leq 1$, convexity and IB (i.e. $w_m \geq 0$) ensure that the right-hand-side of the last expression is non-negative whereas left-hand side is negative, so the inequality holds.

Part 3. Axiom R Effect. Let $w^r(q_F, q_L)$ denote WTP for changed preferences as per Axiom R. Then

$$Q_D^r = -\frac{w^r(q_F, q_L)}{2}q_L$$

for all q_L, and

$$\begin{aligned}\frac{dQ_D^r}{dq_L} &= -\frac{w^r(q_F, q_L)}{2} - \frac{q_L}{2}\frac{dw^r(q_F, q_L)}{dq_L} \\ &= -\frac{w(q_F, q_L)}{2} - \frac{w^r(q_F, q_L) - w(q_F, q_L)}{2} - \frac{q_L}{2}\frac{dw^r(q_F, q_L)}{dq_L}.\end{aligned}$$

From Axiom R the second term is positive and similarly as in part 2 for the sign of the third term.

Part 4. Axiom S Effect. Let $w^s(q_F, q_L)$ denote WTP for changed preferences as in Axiom S**.** Then

$$Q_D^c = -\frac{w^s(q_F, q_L)}{2} q_L$$

is smaller (larger) than Q_D^r if the status quo is smaller (larger) than q_L, and

$$\frac{dQ_D^s}{dq_L} = -\frac{w^s(q_F, q_L)}{2} - \frac{q_L}{2}\frac{dw^s(q_F, q_L)}{dq_L}$$

has an additional positive (negative) term if the status quo is smaller (larger) than q_L.

Q.E.D.

A.12. ***Proof of Proposition 7 (Stackelberg Mini-Game).*** In situation (a) induced preferences $\mathcal{A}^a_{F_s}$ are $\mathcal{A}^a_{F_s^c}$ or $\mathcal{A}^a_{F_s^*}$ depending on whether output x is considered as status quo by the Follower. Axiom S implies $\mathcal{A}^a_{F_s^o}$ **MAT** $\mathcal{A}^a_{F_s^c}$ and $\mathcal{A}^a_{F_s^o}$ **MAT** $\mathcal{A}^a_{F_s^*}$. Similarly, in situation (b) Axiom S implies $\mathcal{A}^b_{F_s^c}$ **MAT** $\mathcal{A}_{F_s^o}$ and $\mathcal{A}^b_{F_s^*}$ **MAT** $\mathcal{A}_{F_s^o}$. By transitivity $\mathcal{A}^b_{F_s}$ **MAT** $\mathcal{A}^a_{F_s}$. Then the the last inequality is straightforward by definition of **MAT** whereas for the first one recall that: (i) q_F^o stays constant (it depends only on s), and NTP along ∂F_s decreases as q_F increases.

Q.E.D.

Appendix B. Instructions

Welcome

This is an experiment about decision-making. You will be paid a $5 participation fee plus an additional positive or zero amount of money determined by the decisions that you and the other participants make, as explained below. Payment is in cash at the end of the experiment. A research foundation has provided the funds for this experiment.

No Talking Allowed

Now that the experiment has begun, we ask that you do not talk. If you have a question, please raise your hand and an experimenter will approach you and answer your question in private.

A Monitor and Two Groups

A monitor will be selected randomly from among those of you who came here today. The rest of you have been divided randomly into two groups, called the First Mover Group and the Second Mover Group.

Complete Privacy

The experiment is structured so that no one — not even the experimenters, the monitor, and the other subjects — will ever know your personal decision in the experiment. You collect your cash payment from a staff person in the Economics Department office who has no other role in the experiment. Your payment is in a sealed envelope with a code letter (A, B, C, etc). Your privacy is guaranteed because neither your name nor your student ID number will appear on any decision records. The only identifying mark on the decision forms will be a code letter known only to you. You will show your code letter to the staff person and nobody else will see it. The experimenters will not be in the department office when you collect you cash payment. This procedure is used to protect your privacy.

The Idea of the Game

The game involves two players, called the First Mover (FM) and the Second Mover (SM), in the roles of producers of an identical good. Each decides how much to produce. The profit for each player is the number of units he decides to produce times price, net of cost. The price of the good decreases as total production increases. If you and the other player produce too much, you will drive down the price and your profits. Of course, if you don't produce much you won't have many units to sell.

To simplify your task, the profits will be calculated for you and shown in an easy-to-read table. Your cash payment will include the profit you earn in one round of the game. The round will be selected randomly at the end of the experiment.

Game Details

Each round the FM chooses between two possible amounts to produce, as shown in a table with two rows. The SM sees the choice of the FM, and then decides among seven possible amounts to produce, as shown in seven columns of the same table. The table shows the profits for both players. The FM's profit is shown in italics in the lower left corner of each box, and the SM's profit is shown in bold in the upper right corner. For example, in Table B.I below, if FM chooses Output=6 and SM then chooses Output=4, then FM's profit is 84 and SM's profit is 56.

TABLE B.I

	SM's Choice of Output Quantity:							
	4	**5**	**6**	**7**	**8**	**9**	**10**	**11**
FM's Output=6	*84* **56**	*78* **65**	*72* **72**	*66* **77**	*60* **80**	*54* **81**	*48* **80**	*42* **77**
FM's Output=9	*99* **44**	*90* **50**	*81* **54**	*72* **56**	*63* **56**	*54* **54**	*45* **50**	*36* **44**

Different Subject Pairs in Every Decision

Each First Mover and each Second Mover will make four decisions. But the pairing of First Movers with Second Movers will be different in every decision. This means that you will interact with a DIFFERENT person in the other group in every decision that you make.

Experiment Procedures and the Monitor

At the beginning of the experiment, the monitor will walk through the room carrying a box containing unmarked, large manila envelopes. Each subject in the First Mover Group will take one of these envelopes from the box. This envelope will contain the experiment decision forms and a code letter.

After the First Movers have made their decisions, they return the experiment decision forms to their large manila envelopes and then walk to the front of the room and deposit the envelopes in the box on the table. It is very important that the First Movers do NOT return their code letters to the large manila envelopes, because they will need them to collect their payoffs.

After all First Movers have deposited their envelopes in the box, the Monitor will take the box to another room in which the experimenters will sort the decision forms and place them in the correct large manila envelopes for the Second Movers. The experimenters will also put code letters in the envelopes for the Second Movers.

Next, the Monitor will walk through the room carrying a box containing unmarked, large manila envelopes. Each subject in the Second Mover Group will take one of these envelopes from the box. This envelope will contain the experiment decision forms and a code letter.

After the Second Movers have made their decisions, they return the experiment decision forms to their large manila envelopes and then walk to the front of the room and deposit the envelopes in the box on the table. It is very important that the Second Movers do NOT return their code letters in the large manila envelopes because they will need them to collect their payoffs.

After all Second Movers have deposited their envelopes in the box, the Monitor will take the box to another room in which the experimenters will record the profits and cash payments determined by the subjects' decisions.

A Roll of a Die Determines Which Decision Pays Money

Although you will make four decisions, only one will pay cash. Which of these decisions will pay cash will be determined by rolling a six-sided die. The experimenter will roll the die in front of you and the monitor will announce which of the numbered sides has ended up on top. The first number from 1 to 4 that ends up on top will determine the page number of the decision that pays cash.

The monitor's cash payment will be the average of all First Movers and Second Movers payments.

Be Careful

Be careful in recording your decisions. If a First Mover forgets to circle one of the rows in the table, or circles both rows on the same decision page, then it will be impossible to ascertain what decision the First Mover made. In that case, the First Mover will get paid 0 and the Second Mover will get paid 60 if that decision page is selected for payoff by the roll of the die. If a Second Mover doesn't circle a column, then it will be impossible to ascertain what decision the Second Mover made. In that case, the Second Mover will get paid 0 and the First Mover will get paid 60 if that decision page is selected for payoff by the roll of the die.

Pay Rates

For each point of profit you earn, the experimenter will put a fixed number of dollars in your envelope. This fixed number is called the pay rate and is written on the board at the front of the room. Today's pay rate is $0.25, which means that every participant earns 25 cents for each profit point shown in the table.

Frequently Asked Questions

Q1: Exactly how are profits calculated in the Tables?

A: Price is 30 minus the sum of FM output and SM output. Marginal cost is 6. Profit is output times (price minus marginal cost). But you don't have to worry about doing the calculation; the Tables do it for you.

Q2: Who will know what decisions I make?

A: Nobody else besides you; that is the point of the private envelopes etc. The experimenters are only interested in knowing the distribution of choices for FMs and SMs, not in the private decisions of individual participants.

Q3: Is this some psychology experiment with an agenda you haven't told us?

A: No. It is an economics experiment. If we do anything deceptive, or don't pay you cash as described, then you can complain to the campus Human

Subjects Committee and we will be in serious trouble. These instructions are on the level and our interest is in seeing the distribution of choices made in complete privacy.

Any More Questions?

If you have any questions, please raise your hand and an experimenter will approach you and answer your question in private. Make sure that you understand the instructions before beginning the experiment; otherwise you could, by mistake, mark a different decision than you intended.

Quiz

(1) In Table B.II below, what are the two possible output choices for the FM? __
(2) Does the SM see the FM's choice? (Y or N)
(3) In Table B.II, can the SM choose:
   (a) Output=5? __(Y or N)
   (b) Output =7?__(Y or N)
   (c) Output=12?__(Y or N)
(4) Suppose the FM chooses the top row (Output = 9) in Table B.II and the SM chooses a middle column (Output = 8).
   (a) How many points will the FM get? __ points
   (b) How much money is that if this is the decision that pays money? $ __
   (c) How much will the SM get in this case? __ points, $ __
(5) In the previous question, if SM chose Output=9 instead of Output=8,
   (a) how many more or fewer points would the SM get? __ more/fewer points
   (b) how many more or fewer points would the FM get? __ more/fewer points
(6) If the FM chooses the top row, what is the maximum number of points that the SM can get? __ the minimum number?__
(7) If the FM chooses the bottom row, what is the maximum number of points that the SM can get? __ the minimum number?
(8) Will the SM ever be able to tell which person made any FM choice? (Y or N)
(9) Will the FM ever be able to tell which person made any SM choice? (Y or N)
(10) Will the experimenter ever be able to tell who made any choices? (Y or N)

TABLE B.II

SM's Choice of Output Quantity:	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
FM's Output=9	*99* **44**	*90* **50**	*81* **54**	*72* **56**	*63* **56**	*54* **54**	*45* **50**	*36* **44**
FM's Output=12	*96* **32**	*84* **35**	*72* **36**	*60* **35**	*48* **32**	*36* **27**	*24* **20**	*12* **11**

REFERENCES

[1] ABBINK, K. , B. IRLENBUSCH, AND E. RENNER (2000): "The Moonlighting Game: An Empirical Study on Reciprocity and Retribution," *Journal of Economic Behavior and Organization,* 42, 265-77.

[2] ANDREONI, J., M. CASTILLO, AND R. PETRIE (2003): "What Do Bargainers' Preferences Look Like? Experiments with a Convex Ultimatum Game," *American Economic Review,* 93, 672-85.

[3] ANDREONI, J., W. HARBAUGH, AND L. VESTERLUND (2003): "The Carrot or the Stick: Rewards, Punishments, and Cooperation," *American Economic Review,* 93, 893-902.

[4] ANDREONI, J. AND J. MILLER (2002): "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica,* 70, 737-53.

[5] BERG, J., J. DICKHAUT, AND K. MCCABE (1995): "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10, 122-42.

[6] BOLTON, G. E. AND A. OCKENFELS (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review,* 90, 166-193.

[7] BOSMAN, R. AND F. VAN WINDEN (2002), "Emotional Hazard in a Power-to-Take Experiment," *Economic Journal,* 112, 147-69.

[8] CHARNESS, G. AND M. RABIN (2002), "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics,* 117, 817-869.

[9] COX, J. C. (2004), "How to Identify Trust and Reciprocity," *Games and Economic Behavior,* 46, 260-281.

[10] COX, J. C., D. FRIEDMAN, AND S. GJERSTAD (2007): "A Tractable Model of Reciprocity and Fairness," *Games and Economic Behavior,* 59, 17-45.

[11] COX, J. C. AND V. SADIRAJ (2007): "On Modeling Voluntary Contributions to Public Goods," *Public Finance Review,* 35, 311-332.

[12] DUFWENBERG, M. AND G. KIRCHSTEIGER (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior,* 47, 268-298.

[13] FALK, A., E. FEHR, AND U. FISCHBACHER (2003). "On the Nature of Fair Behavior," *Economic Inquiry,* 41, 20-26.

[14] FALK, A. AND U. FISCHBACHER (2006), "A Theory of Reciprocity," *Games and Economic Behavior,* 54, 293-315.

[15] FEHR, E. AND S. GAECHTER (2000): "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives,* 14, 159-181.

[16] Fehr, E., G. Kirchsteiger, and A. Riedl (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics,* 108, 437-460.

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

[18] Gale, J., K. Binmore, and L. Samuelson (1995): "Learning to be Imperfect: The Ultimatum Game," *Games and Economic Behavior*, 8, 56-90

[19] Gray, A. (1997): *Modern Differential Geometry of Curves and Surfaces with Mathematica*, 2nd ed. Boca Raton, FL: CRC Press, p. 14-17. see as well, http://mathworld.wolfram.com/Curvature.html.

[20] Guth, W., R. Schmittberger, and B. Schwarze (1982): "An Experimental Analysis of Ultimatum Barga:ining," *Journal of Economic Behavior and Organization,* 4, 367-88.

[21] Hicks, J. (1939). *Value and Capital*. Oxford: Clarendon Press.

[22] Huck, S., W. Muller and H. Normann (2001): "Stackelberg beats Cournot: On collusion and efficiency in experimental markets," *Economic Journal,* 111,749-766.

[23] Hurewicz, W., (1958): *Lectures on Ordinary Differential Equations*. New York: Wiley.

[24] Hurwicz, L. and H. Uzawa (1971): "On the Integrability of Demand Functions," in J.S.Chipman et al. (eds.), *Studies in the Mathematical Foundations of Utility and Demand Theory*. New York: Harcourt, Brace & World.

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

[26] Liebrand, W.B.G. (1984): "The effect of social motives, communication and group sizes on behavior in an n-preson multi stage mixed motive game," *European Journal of Social Psychology,* 14, 239-264.

[27] Protter, M.H. and C.B. Morrey, JR. (1963): *Calculus with Analytical Geometry,* Addison-Wesley Publishing Company, INC.

[28] Rockafellar, R. T. (1970): *Convex Analysis*, Princeton University Press.

[29] Samuelson, P. (1947): *Foundations of Economic Analysis*. Cambridge, Mass., Harvard University Press.

[30] Simon, C.P. and L. Blume, (1994): *Mathematics for Economists*, W.W. Norton & Company, Inc.

[31] Smith, A. <1759> (1976): *The Theory of Moral Sentiments*. Indianapolis, Liberty Classics.

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

[33] Sonnemans, J, F. van Dijk, and F. van Winden (2005): "On the Dynamics of Social Ties Structures in Groups," *Journal of Economic Psychology,* forthcoming.

[34] Varian, H. R. (1992): *Microeconomic Analysis*, third ed., Norton.

[35] Varian, H. R. (1994): "Sequential provision of public goods," *Journal of Public Economics,* 53, 165-186.

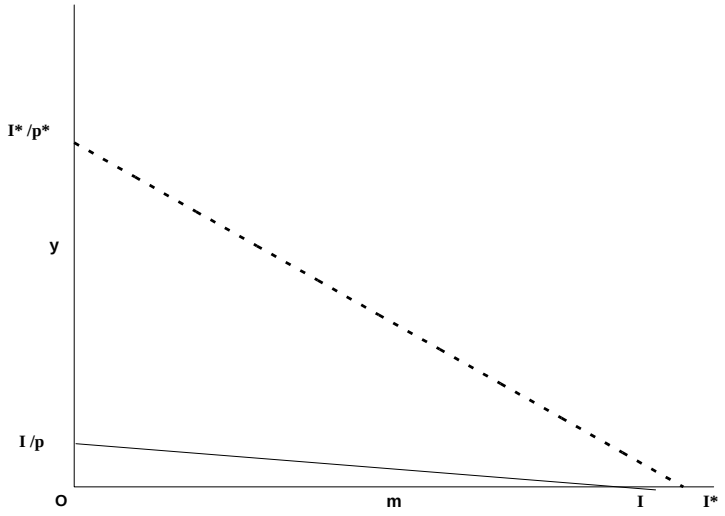

FIGURE 1. Standard Budget Set

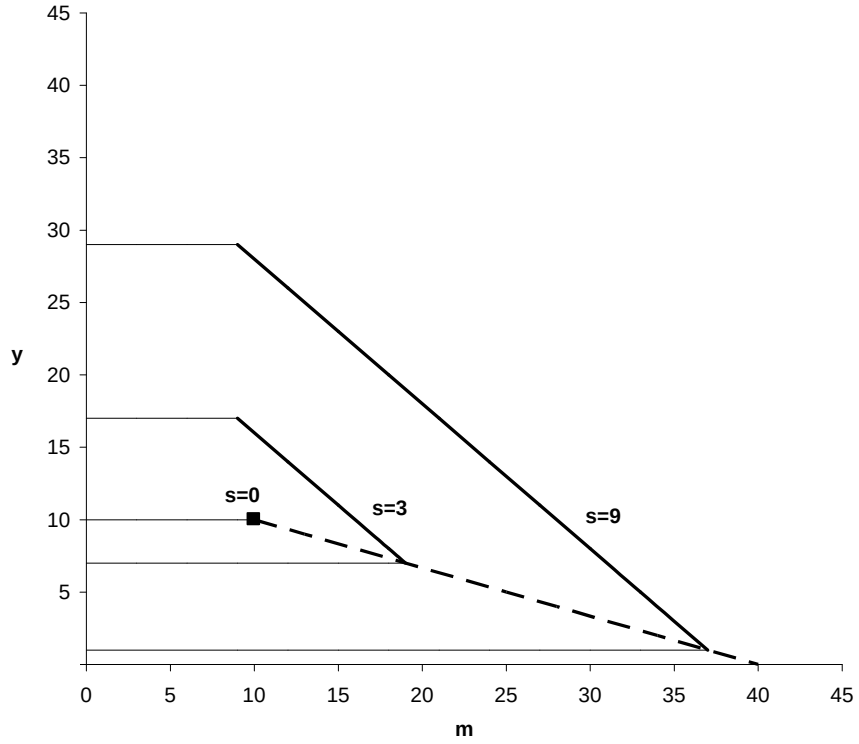

FIGURE 2. Investment Game, Second Mover's Opportunity Set.

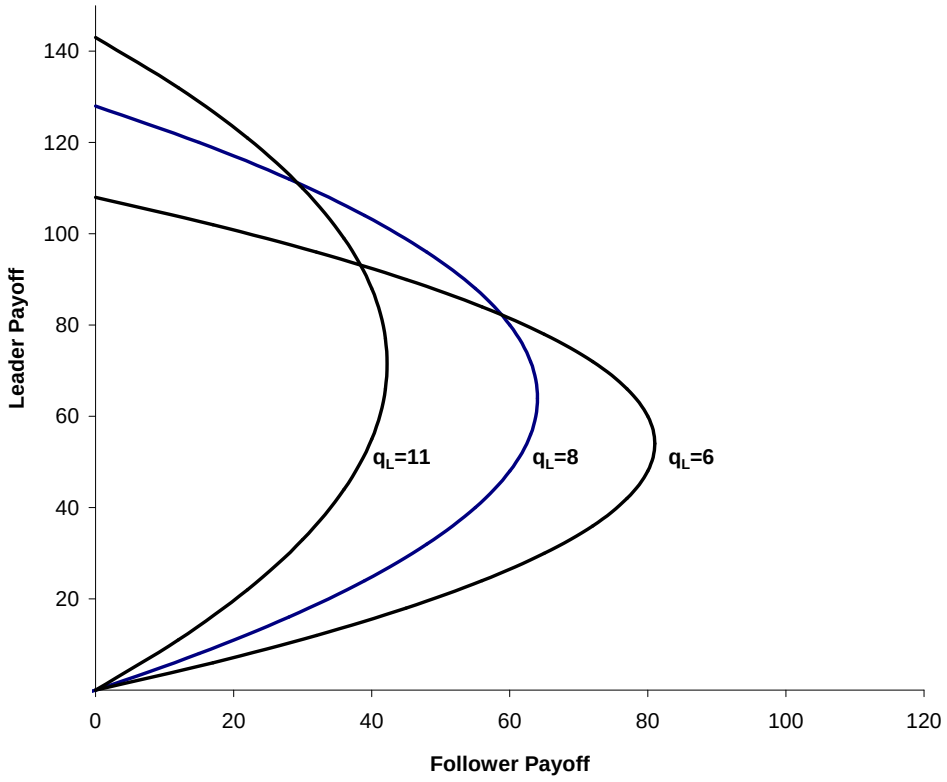

FIGURE 3. Stackelberg Duopoly Game, Follower's Opportunity Set.

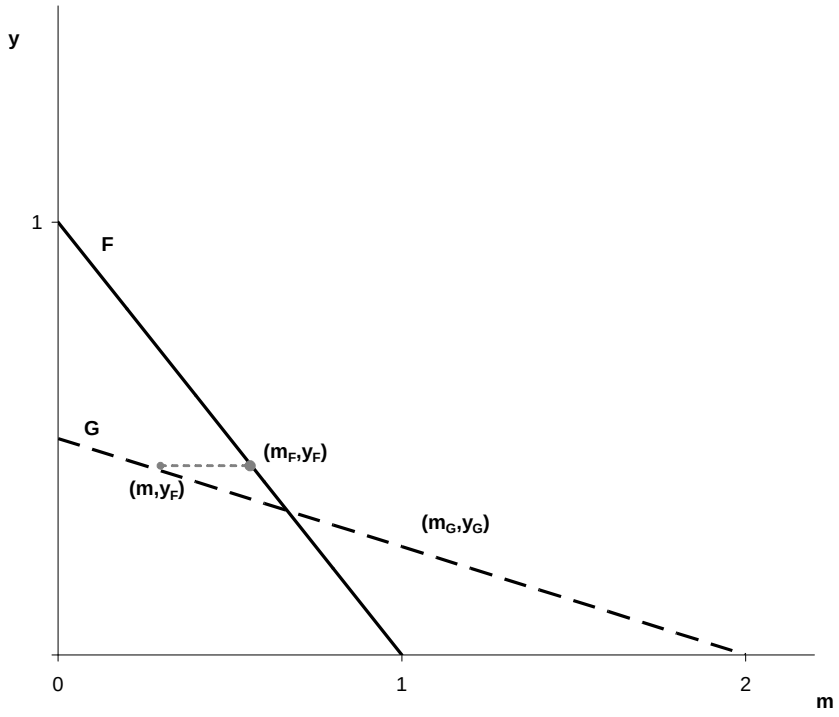

FIGURE 4. Illustration of Example 5.1.

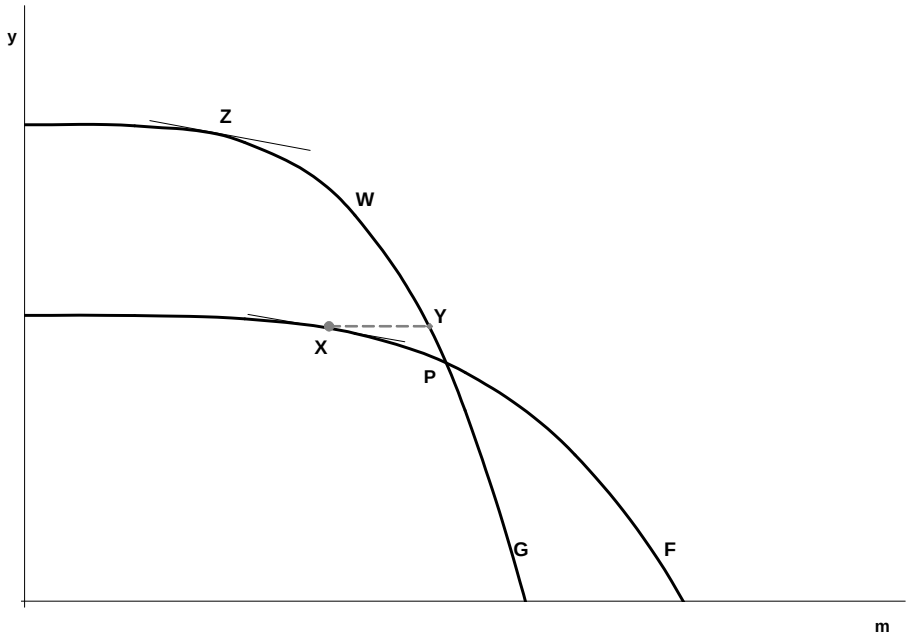

FIGURE 5. Parts 2 and 3 of Proposition 3 predict that, with unchanged IB preferences, the choice W on the Eastern boundary of G will lie between points Y and Z. Part 1 of the proposition predicts that W is north of point P.

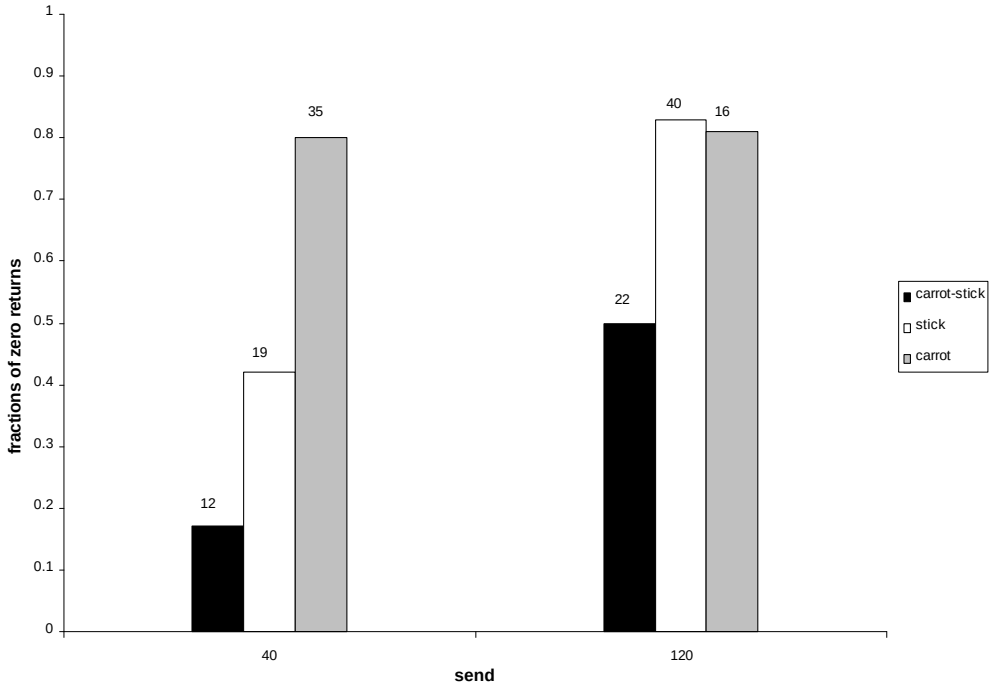

FIGURE 6. Fractions of zero returns across games when the first movers send 40 or 120. Only data from the last five rounds are included.