

NOTES AND COMMENTS

GIVING ACCORDING TO GARP: AN EXPERIMENTAL TEST OF THE CONSISTENCY OF PREFERENCES FOR ALTRUISM

BY JAMES ANDREONI AND JOHN MILLER¹

1. INTRODUCTION

SUBJECTS IN ECONOMIC LABORATORY experiments have clearly expressed an interest in behaving unselfishly. They cooperate in prisoners' dilemma games, they give to public goods, and they leave money on the table when bargaining. While some are tempted to call this behavior irrational, economists should ask if this unselfish and altruistic behavior is indeed self-interested. That is, can subjects' concerns for altruism or fairness be expressed in the economists' language of a well-behaved preference ordering? If so, then behavior is consistent and meets our definition of rationality.

This paper explores this question by applying the axioms of revealed preference to the altruistic actions of subjects. If subjects adhere to these axioms, such as GARP, then we can infer that a continuous, convex, and monotonic utility function could have generated their choices. This means that an economic model is sufficient to understand the data and that, in fact, altruism is rational.

We do this by offering subjects several opportunities to share a surplus with another anonymous subject. However, the costs of sharing and the surplus available vary across decisions. This price and income variation creates budgets for altruistic activity that allow us to test for an underlying preference ordering.

We found that subjects exhibit a significant degree of rationally altruistic behavior. Over 98% of our subjects made choices that are consistent with utility maximization. Only a quarter of subjects are selfish money-maximizers, and the rest show varying degrees of altruism. Perhaps most strikingly, almost half of the subjects exhibited behavior that is exactly consistent with one of three standard CES utility functions: perfectly selfish, perfect substitutes, or Leontief. Those with Leontief preferences are always dividing the surplus equally, while those with perfect substitutes preferences give everything away when the price of giving is less than one, but keep everything when the price of giving is greater than one. Using the data on choices, we estimated a population of utility functions and applied these to predict the results of other studies. We found that our results could successfully characterize the outcomes of other studies, indicating still further that altruism can be captured in an economic model.

¹ We are grateful to Ted Bergstrom, Mahmoud El-Gamal, Bill Harbaugh, Glenn Harrison, Matthew Rabin, Larry Samuelson, Perry Shapiro, and Hal Varian for their helpful comments, Peter Brady and Isaac Rischall for expert research assistance, and Lise Vesterlund for help collecting the data. We also owe a great debt to an editor and to three anonymous referees for extremely helpful remarks. For financial support, Andreoni acknowledges the National Science Foundation and Miller thanks Carnegie Mellon University.

We also addressed one further puzzle from experiments: Are altruistic preferences monotonic? Evidence suggests that some subjects are willing to sacrifice a portion of their own payoff to reduce the payoff of another. That is, preferences may not be monotonic, but instead may show jealousy or spite. We tested this by presenting subjects a series of upward-sloping but finite budgets. We found that, in fact, a sizable minority of subjects, 23%, have preferences that, while convex, are not monotonic.

We conclude that, indeed, subjects exhibit a consistent preference for altruism. When altruism is rephrased in the language of prices and income, then we uncover preferences that are predictable and well-behaved. In the next section we present a formal theoretical framework for our study. In Section 3 we outline the revealed preference analysis. In Section 4 we present the experimental design. Sections 5, 6, and 7 present results and analysis. Section 8 explores how well our approach can predict behavior from outside our sample. Section 9 addresses the monotonicity of preferences. Section 10 is a conclusion.

2. TEMPLATE FOR ANALYSIS

We begin by looking at a nonstrategic environment. This is a natural first step, since we should first confirm that preferences are convex in a fixed environment. Once this has been established, then we can begin the more intensive study of how strategy spaces, payoff possibilities, intentions, social cues, and other environmental changes can shift and mold preferences.

Let π_i represent monetary payoffs for person i and let Π be the set of possible payoffs for a game. For simplicity, consider choices made by person s , for self, that have consequences for his own payoff, π_s , and the payoff of one other person, π_o . Any choice in the strategy space for person s implies a mapping into the set of payoffs. Hence, for a particular (nonstrategic) setting, person s can be thought of as choosing the $(\pi_s, \pi_o) \in \Pi$ that maximizes utility. If we assume that subjects in experiments are money maximizers, then we are assuming that they maximize utility of the form $U_s = \pi_s$. To capture the possibility for altruism, however, we must allow a more general form of utility,

$$(1) \quad U_s = u_s(\pi_s, \pi_o).$$

Given that subjects actually make choices over the variables (π_s, π_o) , it seems natural to check first for convex preferences in this space.²

How do we envision a more general model that applies to more complex and changing environments? Let γ be a vector of attributes of a game. This could include the specific economic variables like rules of the game, as well as social variables like the level of anonymity, the sex of one's opponent, or the framing of the decision, all of which are known to affect the outcome. Future work will have to explore the more general assumptions that for a given γ the preferences $U_s = u_s(\pi_s, \pi_o; \gamma)$ are well-behaved with respect to (π_s, π_o) and that these preferences shift systematically as γ changes.

3. REVEALED PREFERENCE AXIOMS

Let A, B, \dots, Z be distinct bundles of alternatives, each lying on a linear budget constraint. Then define two concepts (see Varian (1993)):

² Note that π represents a change in consumption, not consumption *per se*. Our approach does not preclude an assumption that individuals have preferences over total consumption. Obviously, if preferences over total consumption are well-behaved, then preferences over π_s and π_o will be as well.

DIRECTLY REVEALED PREFERRED: A is *directly revealed preferred* to B if B was in the choice set when A was chosen.

INDIRECTLY REVEALED PREFERRED: If A is directly revealed preferred to B , B is directly revealed preferred to C , . . . to Y , and Y is directly revealed preferred to Z , then A is *indirectly revealed preferred* to Z .

The classic revealed preference axioms are due to Samuelson (1938) and Houthakker (1950):

WEAK AXIOM OF REVEALED PREFERENCE (WARP): *If A is directly revealed preferred to B , then B is not directly revealed preferred to A .*

STRONG AXIOM OF REVEALED PREFERENCE (SARP): *If A is indirectly revealed preferred to B , then B is not directly revealed preferred to A .*

WARP is necessary and SARP is both necessary and sufficient for the existence of strictly convex preferences that could have produced the data. Varian (1982), in applying the theorems of Afriat (1967), generalized the theory to allow indifference curves that are not strictly convex:

GENERALIZED AXIOM OF REVEALED PREFERENCE (GARP): *If A is indirectly revealed preferred to B , then B is not strictly directly revealed preferred to A , that is, A is not strictly within the budget set when B is chosen.*

Note that if choices violate WARP they must also violate SARP, and if they violate GARP then they must also violate SARP, but the opposite is not true. As Varian shows, satisfying GARP is both a necessary and sufficient condition for the existence of well-behaved preferences, given linear budget constraints.

4. EXPERIMENTAL DESIGN

We will employ a modified version of the Dictator Game. In the original dictator game, developed by Forsythe, et al. (1994), subjects divide m dollars between themselves and another subject so that $\pi_s + \pi_o = m$. In our experiment, each subject is given a menu of choices with different endowments and prices for payoffs, for instance $\pi_s + p\pi_o = m$. These budget sets over payoffs cross in ways that provide a test for whether well-behaved preferences of the form $u_s(\pi_s, \pi_o)$ could explain the data.

Specifically, the experiment was conducted with volunteers from intermediate and upper-level economics courses. There were 5 experimental sessions of 34 to 38 subjects each, for a total of 176 subjects. Each subject's task was to allocate "tokens" under a series of different budgets. In sessions 1–4 there were eight budget choices, while session 5 offered 11 budgets. As we discuss later, session 5 was added last to test the strength of the results from sessions 1–4.

Each of the decision problems differed in the number of tokens to be divided and the number of points a token was worth to each subject. Tokens were worth either 1, 2, 3, or 4 points each. The total number of tokens available was either 40, 60, 75, 80, or 100. Subjects made their decision by filling in the blanks in a statement like, "Divide 60 tokens: *Hold* ____ at 1 point each, and *Pass* ____ at 2 points each." Subjects were encouraged to

TABLE I
ALLOCATION CHOICES

Budget	Token Endowment	Hold Value	Pass Value	Relative Price of Giving	Average Tokens Passed
1	40	3	1	3	8.0
2	40	1	3	0.33	12.8
3	60	2	1	2	12.7
4	60	1	2	0.5	19.4
5	75	2	1	2	15.5
6	75	1	2	0.5	22.7
7	60	1	1	1	14.6
8	100	1	1	1	23.0
9 ^a	80	1	1	1	13.5
10 ^a	40	4	1	4	3.4
11 ^a	40	1	4	0.25	14.8

^aWere only used in session 5, others used in all sessions.

use a calculator to check their decisions. The decision problems were presented in random order to each subject. Subjects were told that the experimenter would choose one of the decision problems at random and carry it out with another randomly chosen subject as the recipient. Finally, subjects were told that each point earned would be worth \$0.10 in payoff, hence 75 points would earn \$7.50. The budgets offered are shown in Table I.

Notice that each allocation decision presents a convex budget set. Consider budget 1. Here transferring one token raises the other subject's payoff by 1 point, and reduces one's own payoff by 3, implying that the price of the opponent's payoff is 1 and the price of self-payoff is 0.33. Hence, the token endowment is an income variable, the inverse of the hold value is the price of self-payoff π_s , and the inverse of the pass value is the price of other payoff π_o . When the relative price is 1, as in budgets 7, 8, and 9, the choices are like standard dictator games.

We conducted each session by assembling subjects in a large classroom. We distributed envelopes containing a copy of the instructions, a pencil, an electronic calculator, and a "claim check" that subjects used to claim their earnings.³

In session 5, in addition to the three new budgets listed in Table I, the subjects made five additional decisions. We call this part 2. Here subjects were assigned allocations of tokens, but were asked to decide how many cents each token would be worth, from 0 to 10 cents each. For example, subjects filled out questions like this:

Divide 140 tokens: *Hold* 10 at 1 point each, and *Pass* 130 at 1 point each. How many cents should each point be worth? (circle one) 0 1 2 3 4 5 6 7 8 9 10

³The instructions were read aloud by the experimenter. Subjects then filled out the experimental questionnaires, and returned them to the blank envelopes. The envelopes were collected, shuffled, and taken to a neighboring room. Payments for each subject were calculated and put into an envelope labeled with the subject's number. The payment envelopes were then brought back to the room with the waiting subjects. An assistant who had remained in the room with the subjects, and hence had no knowledge of what may be in the payoff envelopes, asked subjects to present their claim checks, one at a time, and gave them their payment envelopes. Since we calculated payoffs in a room away from the subjects, we also used a monitor, selected randomly from among the subjects, to verify to other subjects that the promised procedures for calculating payoffs were followed. Sessions 1–4 lasted less than one hour and subjects earned an average of \$9.60. Session 5 lasted about 70 minutes, and subjects earned an average of \$19.74.

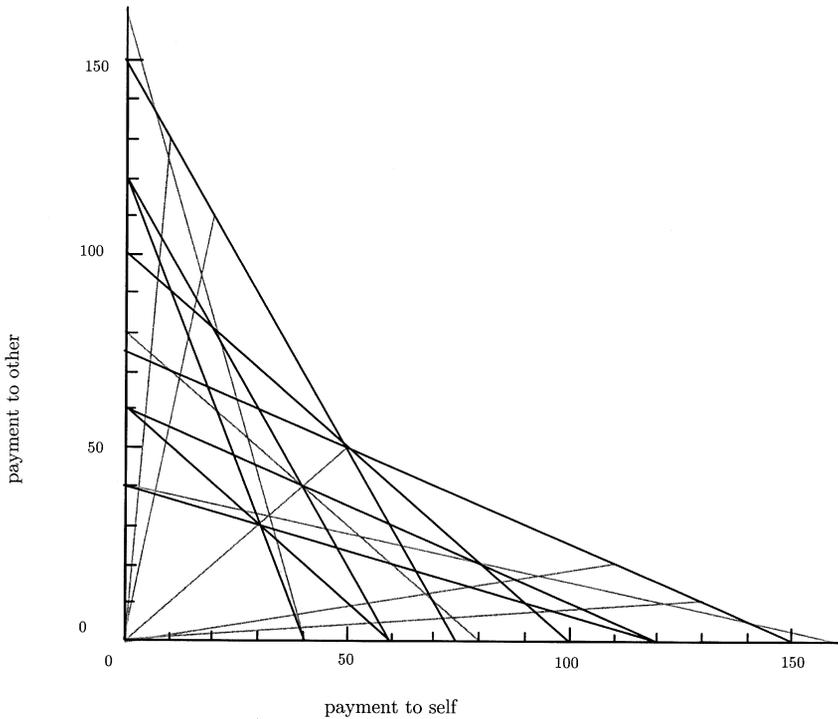


FIGURE 1.—Budget constraints offered subjects.

The five choices, presented in random order across subjects, had assignments of hold and pass quantities of (10, 130), (20, 110), (50, 50), (110, 20), and (130, 10), and all tokens were worth 1 point each in every decision. One of the five decisions was chosen at random to be carried out.

Notice that these choices are equivalent to giving subjects budget constraints that slope *up*. This will allow us to test the conjecture that preferences are perhaps nonmonotonic, and to see if there is some “rational jealousy.” For instance, in the example given above, if the subject values points at 10 cents each, then she will earn \$1 and her opponent will earn \$13. If this inequality is displeasing to the subject, she may value points at, say, 6 cents each, in which case she will earn \$0.60 and her opponent will earn \$7.80. At the extreme she could value points at 0, in which case both subjects earn nothing.

The full menu of budgets offered is shown in Figure 1. Those presented in just session 5 are in grey.⁴

5. CHECKING RATIONALITY

We begin by looking at the downward sloping budgets. The average choices across the 11 budgets are shown in Table I, where all subjects saw budgets 1–8, and only session 5 saw the additional budgets 9–11.

⁴ A copy of the instructions used in the experiment is available from the authors, or at www.ssc.wisc.edu/~andreoni/.

TABLE II
 VIOLATIONS OF REVEALED PREFERENCE

	Subject	Number of Violations			Critical Cost Effic. Index
		WARP	SARP	GARP	
Sessions 1-4:	3	1	3	2	1*
	38	2	7	7	0.92
	40	3	8	7	0.83
	41	1	1	1	1*
	47	1	1	1	1*
	61	1	4	3	0.91
	72	1	1	1	1*
	87	1	1	1	1*
	90	1	1	1	0.98
	104	1	2	1	1*
	126	1	3	1	1*
	137	1	1	1	1*
	139	1	1	1	1*
	Session 5:	211	1	2	2
218		1	2	1	1*
221		1	1	1	1*
223		1	1	1	1*
234		1	1	1	1*

*Indicates that an ϵ -change in choices eliminates all GARP violations.

First, are the data representative of other studies? To answer this we look at those with slopes of minus one, budgets 7-9. With the pie of six dollars, an average of \$1.46, or 24.5 percent of the pie, is given away. With the ten dollar pie, an average of \$2.30 (23.0 percent) is given away, and for the eight dollar pie, \$1.35 (16.9 percent) is given. Combining the three, our subjects gave away 23 percent of their budget when the relative price was one. This is strikingly similar to Forsythe, et al. (1994) who found 22.2 percent of a five dollar pie and 23.3 percent of a ten dollar pie were given away.

Second, did the subjects choose rationally, and if they had violations of the revealed preference axioms, how severe were they? One measure of the severity is Afriat's (1972) Critical Cost Efficiency Index (CCEI). Roughly speaking, the CCEI gives the amount we would have to relax each budget constraint in order to avoid violations.⁵ The closer the CCEI is to one, the smaller we would have to shrink any budgets to avoid violations. Note that it is possible for the CCEI to be equal to 1 when moving one choice by an infinitesimal amount would remove the violation.⁶ Since there is no natural significance threshold for the CCEI, we follow Varian's (1991) suggestion of a threshold of 0.95.

⁵ Define a generalization of the revealed preference relation $R^D(e')$ such that $x' R^D(e') x$ iff $e' p' x' \geq p' x$, that is, x would not be affordable at a fraction e' of the income available when the person chose x' . Define $R(e')$ as the transitive closure of $R^D(e')$. Then define $\text{GARP}(e')$ as "if $x' R(e') x^s$, then $e' p^s x^s \leq p^s x'$." Then the CCEI is the highest value of e' such that there are no violations of $\text{GARP}(e')$. See Varian (1991).

⁶ This will happen when, for instance, choice A was on the budget line when B was chosen, but B was strictly within the budget when A was chosen. In addition, we adopt as a convention that if two bundles are directly revealed preferred to each other, this counts as one violation of WARP, not two.

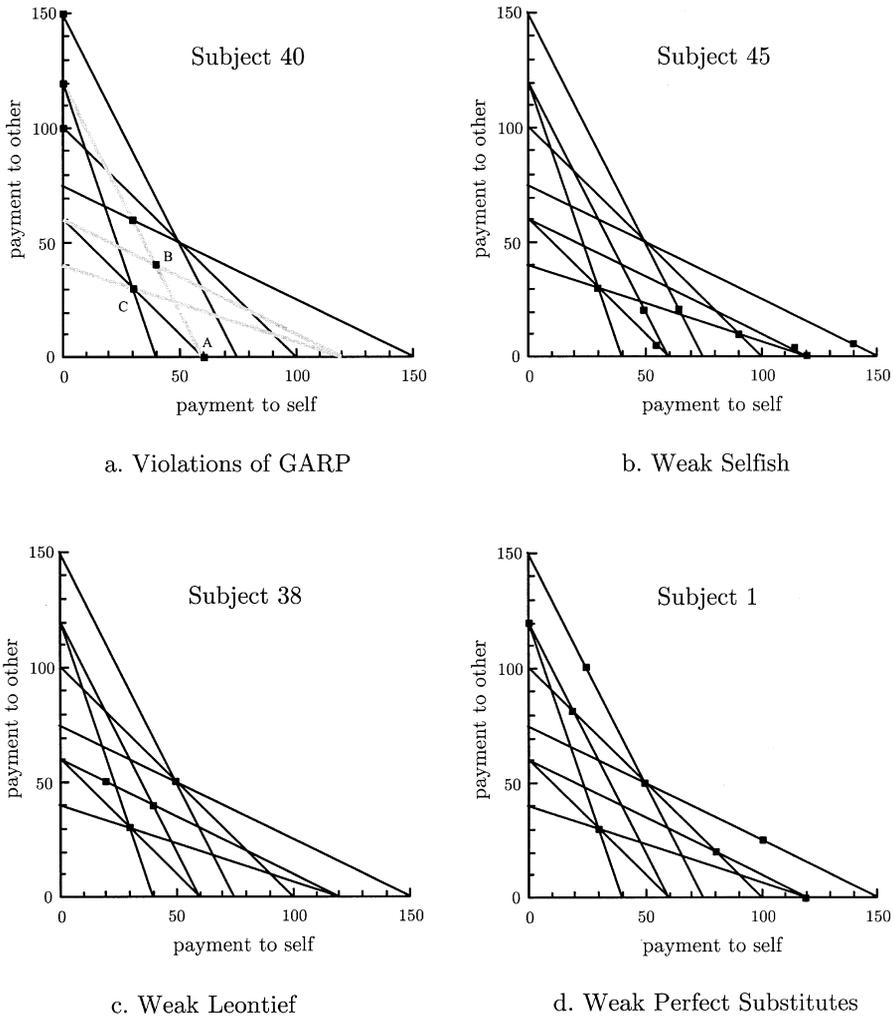


FIGURE 2.—Analyzing individual preferences.

The violations of revealed preference are listed in Table II. Of the 176 subjects, 18 of them violated one or more of the revealed preference axioms. Of these, 4 had violations CCEI indices of less than 1, and three of those were below the 0.95 threshold.

The choices of subject 40, the subject with the most severe violations, are shown in Figure 2(a). It is easy to spot violations of all three notions of revealed preference here. Consider three allocations, labelled *A*, *B*, and *C* on the shaded budget constraints. Allocation *A* is revealed preferred to *C*, and *C* to *A*, violating WARP. *C* is indirectly revealed preferred to *B*, but *B* is strictly directly revealed preferred to *C*, violating SARP and GARP. Small shifts along these budgets would not diminish these violations. Hence, there is no well-behaved preference ordering that could have generated the choices of subject 40.

With the exception of 3 subjects (1.7 percent), we see that behavior can indeed be rationalized by a quasiconcave utility function.⁷ This raises the third question—how stringent is our revealed preference test? The test will be stronger the more opportunities it gives subjects to make choices that violate the axioms. Bronars (1987) designed a test that looks at this question from an *ex ante* perspective. The test can be described as finding the probability that a person whose behavior on any budget was purely random would violate GARP. In particular, artificially generate choices by randomly drawing points on each budget line using a uniform distribution across the entire line. Then ask, what is the probability that such an exercise will lead to a violation of revealed preference?

Bronars' test has been applied several times to experimental data. Cox (1997) considered three consumption goods and seven budgets. His study has a Bronars power of 0.49 (49% of random subjects had violations), and supported rational choice. Sippel (1997) conducted two experiments with 8 goods, ranging from Coca-Cola to video games, and 10 budgets. He used Bronars powers of 0.61 and 0.97, but found 95% of subjects violated GARP. Mattei (2000), in a study similar to Sippel's, considered 8 goods and 20 budgets, with Bronars power of 0.99. He found violations in 30–50% of subjects. Harbaugh, Krause, and Berry (2001), in a study of children, considered 2 goods—chips and juice boxes—and 11 budgets and found that the random subjects violated GARP an expected 8.9 times. Harbaugh, et al., found that students from sixth grade and above were largely consistent with GARP.

We first conducted Bronars' test on the eight budgets of sessions 1–4. Generating a random population of 50,000 subjects, we found 78.1 percent of the random subjects violated all three axioms, with an average of 2.52 violations of WARP, 7.68 of SARP, and 7.52 of GARP. We repeated the analysis using the 11 budgets of session 5 and found that the power increased to 94.7 percent of the random population violating the axioms. There was an average of 4.39 violations of WARP, 17.62 of SARP, and 17.28 of GARP.

Another way to look at the power of the revealed preference test is from an *ex post* perspective. For instance, if all subjects chose only corner solutions, then the selected budgets would not yield much information about the rationality of choices, regardless of Bronars' power test. Hence, we designed a power test by bootstrapping from the sample of subjects. In particular, we created a population of 50,000 synthetic subjects in which the choices on each budget were randomly drawn from the set of those actually made by our subjects. With this test, for session 1–4 we found 76.4 percent of the synthetic subjects had violations, averaging 2.3 of WARP, 7.43 of SARP, and 6.5 of GARP. For session 5, we found 85.7 percent of the subjects had violations, averaging 3.14 of WARP, 10.60 of SARP, and 9.61 of GARP.

6. INDIVIDUAL PREFERENCES

Given that subjects' behavior is rationalizable, we can now try to determine the form of utility functions. Looking at the individual data, we found that a large fraction of the subjects could be fit with a well-known utility function. First, 40 subjects, about 22.7 percent, behaved perfectly selfishly, hence $U(\pi_s, \pi_o) = \pi_s$ could rationalize these data. Second, 25 subjects or 14.2 percent, provided both participants with exactly equal payoffs, implying

⁷ Note that in some cases we cannot preclude concave preferences. This is true for subjects who choose only at corner solutions.

TABLE III
SUBJECT CLASSIFICATION BY PROTOTYPICAL UTILITY FUNCTION

Utility Function	Fit		Total
	Strong	Weak	
Selfish	40	43	83 (47.2%)
Leontief	25	28.5 ^a	53.5 (30.4%)
Perfect Substitutes	11	28.5 ^a	39.5 (22.4%)

^aOne subject was equidistant from strong Leontief and Substitutes.

Leontief preferences of $U(\pi_s, \pi_o) = \min\{\pi_s, \pi_o\}$. Finally, 11 subjects, 6.2 percent, allocated their tokens to the person with the highest redemption value (the lowest price), suggesting $U(\pi_s, \pi_o) = \pi_s + \pi_o$, that is, preferences of perfect substitutes.⁸

These three groupings account for 43 percent of the subjects. This led us to find a means for clustering the remaining subjects by similarities in their choices. We tried several options, but all led to similar classifications of subjects.⁹ Table III lists the simplest of these classifications, which clusters subjects into groups that minimize the distance to choices from one of the three utility functions just described. Hence, we refer to the three inexact classifications as weaker forms of the first three. For illustration, Figures 2b, 2c and 2d show examples of subjects who fit the weak categories.

The finding of six main types of preferences is striking for two reasons. First, these categories show consistency within each subject—43 percent of subjects fit a standard utility function exactly. Second, and perhaps more importantly, there is a great deal of heterogeneity across subjects. People differ on whether they care about fairness at all, and when they do care about fairness the notion of fairness they employ differs widely, ranging from Rawlsian (Leontief) to Utilitarian (perfect substitutes). Clearly this heterogeneity of preferences is important and will have to be captured by any theory of fairness and altruism.

7. ESTIMATING PREFERENCES

This section puts more structure on the preferences of the 57 percent of subjects in the weak categories of the prior section. If we were to characterize the preferences of these subjects, what functions would best capture their behavior?

In estimating utility functions, we must first determine the number of unique utility functions to estimate. Since we have eight to eleven observations on each subject we could, in principle, estimate unique utility functions for each individual. For sake of parsimony, however, we opt instead to pool subjects into groups based on the criteria used to generate Table III.¹⁰ To the extent that this is inaccurate it will dilute the precision of our prediction.

⁸ Among these, there is variance in their choices in the case where the self and other prices were equal. Three of the eleven subjects divided tokens evenly, while six kept all the tokens. One divided evenly when the pie was six dollars, but kept the whole pie when it was ten dollars. A final subject gave all the pie to the other subject on both allocation decisions.

⁹ We also used Bayesian algorithms, adaptive search routines, and minimization of within-group variance.

¹⁰ We are assuming that subjects in the three “strong” categories made choices that were measured without error, hence their utility functions are known. This is clearly a simplifying assumption, since, for instance, a person we call perfectly selfish may show elasticity to demands if we examined a wider

TABLE IV
ESTIMATES OF PARAMETERS (STANDARD ERRORS) FOR CES UTILITY
FUNCTIONS FOR THE THREE WEAK TYPES

	Weak Selfish	Weak Leontief	Weak Perf. Subst.
$A = [a/(1-a)]^{1/(1-\rho)}$	20.183 (5.586)	1.6023 (0.081)	2.536 (0.311)
$r = -\rho/(1-\rho)$	-1.636 (0.265)	0.259 (0.067)	-2.022 (0.188)
a	0.758	0.654	0.576
ρ	0.621	-0.350	0.669
σ	-2.636	-0.741	-3.022
s.e.-self	0.2216 (0.011)	0.179 (0.009)	0.244 (0.014)
ln likelihood	-107.620	52.117	-69.583
Number of cases	380	230	242

Next we must address the question of what functional form to estimate. We considered three different approaches: Cobb-Douglas, Linear-Expenditures Model, and Constant Elasticity of Substitution (CES). Of these, the CES is the most appealing. First, it provides the best fit, across a number of measures, for all three weak types. Second, all the preferences of all six types of subjects can be described with different parameters of the same utility function; hence differences are easily interpreted with an economic rationale. For brevity, therefore, we report the results only for the CES utility function.¹¹

The CES utility function can be written $U_s = (a\pi_s^\rho + (1-a)\pi_o^\rho)^{1/\rho}$. The share parameter a indicates selfishness; ρ captures the convexity of preferences through the elasticity of substitution, $\sigma = 1/(\rho - 1)$. Before solving for demands, normalize budgets by choosing self-payoff to be the numeraire, so $\pi_s + (p_o/p_s)\pi_o = m/p_s$, or simply $\pi_s + p\pi_o = m'$. Maximizing yields the demand function

$$\begin{aligned}\pi_s(p, m') &= \frac{[a/(1-a)]^{1/(1-\rho)}}{p^{-\rho/(\rho-1)} + [a/(1-a)]^{1/(1-\rho)}} m' \\ &= \frac{A}{p^r + A} m',\end{aligned}$$

where $r = -\rho/(1-\rho)$ and $A = [a/(1-a)]^{1/(1-\rho)}$.

Since subjects' choices are censored at both ends of the budget constraint, we estimated the parameters r and A for each weak type using two-limit tobit maximum likelihood, with the restriction that $0 \leq \pi_s/m' \leq 1$. We also found that the error term was heteroskedastic when demands were specified in levels. Hence, to assure homoskedasticity, demands were estimated as budget shares with an i.i.d. error term.

The results of the estimation are shown in Table IV where the decisions of each subject are pooled for each category of subject. The estimated parameters r and A are all signifi-

range of prices. This may weaken the predictive power of our approach, especially when considering prices outside the range employed in the experiment.

¹¹ The results of the more complete analysis are available from the authors.

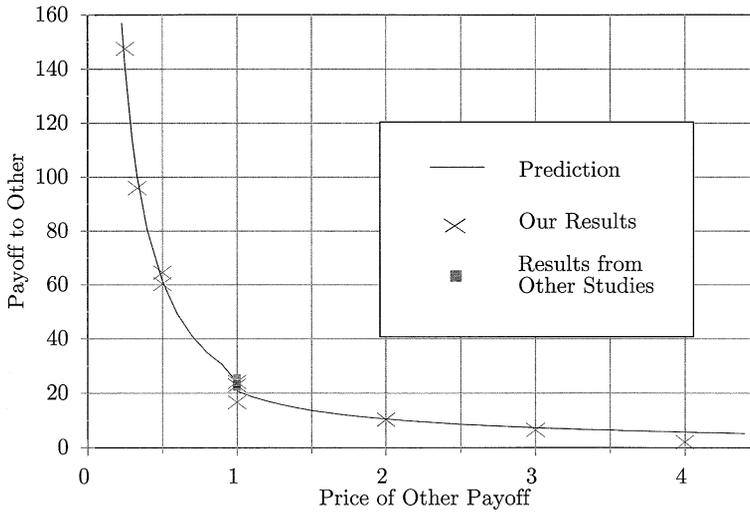


FIGURE 3.—Payment by dictator with endowment of 100.

cant beyond the 0.001 level for all three categories. We also report s.e.-self, the estimated standard error for the residual in the estimation equation for payoff to self. This parameter is important for predicting the distribution of choices from these utility functions.

Using the estimates of A and r to solve for a , ρ , and σ , Table IV shows some interesting differences across types. First, the share parameter a differs substantially, with weakly selfish having the highest and most selfish value. We also see that the elasticity of substitution for the weak Leontief utility function is $\sigma = -0.74$, showing a strong complementarity between π_s and π_o . The elasticities of substitution for the weak selfish is $\sigma = -2.63$ and for weak perfect substitutes is $\sigma = -3.02$, indicating both have very flat indifference curves, but those for the weakly perfect substitutes are slightly flatter.

8. PREDICTION

In this section we explore whether our findings are consistent with behavior in other experiments with similar incentives. Look first at dictator games. Figure 3 illustrates our prediction for a dictator game in which people allocate a pie of 100 with a variable price, that is $\pi_s + p\pi_o = 100$. We do this by using the three estimated utility functions and the three exact utility functions to predict choices of subjects.¹² We then apply a weight to each of the six predicted values based on their frequency reported in Table III. This gives us an overall prediction for average choice at a given price. Along with the prediction, we also plot our data and five results from four other studies.¹³ Note that there is a high level

¹² A technical appendix is available from the authors, or at www.ssc.wisc.edu/~andreoni/.

¹³ These are Forsythe, et al. (1994), Cason and Mui (1997), and Bohnet and Frey (1999a, b). Two prominent studies not included are Hoffman, McCabe and Smith (1996), and Eckel and Grossman (1996). These both employ a “double blind” procedure that has seemed to alter the environment significantly from those our study is meant to capture. These two found average giving of 9.2 percent and 15 percent, respectively, in the double-blind environment.

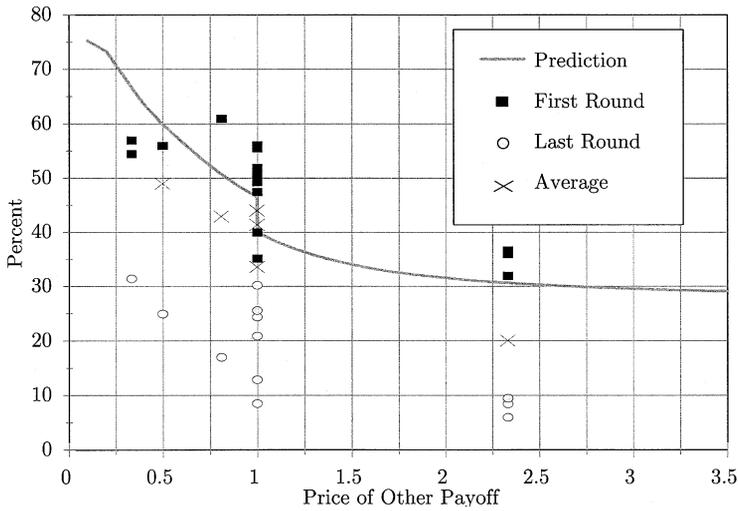


FIGURE 4.—Percent of endowment in the public good. Predicted and results from other studies.

of overall accuracy of our estimating strategy at fitting our data, and consistency with the observations from other experiments.

Are there other experimental games with more price variability that we can use to evaluate and apply our predictions? One related, but imperfect, setting is the linear public goods game. In this game, a person is given a budget of tokens that can be spent on either the private good or the public good. Tokens spent on the private good earn one cent each, while tokens spent on the public good earn α cents ($0 \leq \alpha \leq 1$) for all subjects. Thus, a person can transfer payoff to other subjects at a rate of $(1 - \alpha)/\alpha$. That is, linear public goods games are multi-person dictator games with a price $p = (1 - \alpha)/\alpha$.

The linear public goods game is an imperfect application for several reasons. First, it is often repeated, hence allowing learning. This suggests looking at the first round, since there is no experience. However, since forward-looking subjects may play strategically, we may instead want to look at only the final round. Second, public goods games typically have from four to 100 subjects, whereas our estimates were based on two-person games. We can partially address this problem by considering only small groups of four or five subjects. Given these differences, therefore, any comparison with our data and public goods games will be only suggestive.

The estimated giving curve in public goods games is shown in Figure 4, where we assume subjects care about the per-capita transfer. We also show the results from several public goods experiments, including the first round, last round, and the average across rounds.¹⁴ While there is a wide degree of variance in the outside results, the demand curve generated from our data is quite suggestive of an underlying behavioral regularity

¹⁴ These are Isaac and Walker (1988), Andreoni (1988, 1995a, 1995b), Weimann (1994), Fisher, Isaac, Schatzberg, and Walker (1995), Laury, Walker, and Williams (1995), Croson (1996), and Sonnemans, Schram, and Offerman (1999). Another prominent study not shown separately is by Isaac, Walker, and Williams (1994), which uses the small group data from the Isaac and Walker (1988) study.

for the first round and the average. For the final round, where learning and experience have taken place but in which no strategic play is possible, the data appear to be shifted down in a somewhat parallel fashion.

A third place to apply these results is to prisoner’s dilemma games. Andreoni and Miller (1993) published a study in which subjects participated in 200 rounds of prisoner’s dilemma, with randomly assigned partners. Cooperation in this game averaged 20 percent. Given the payoff parameters and the likelihood of meeting a cooperator, we find that the strong perfect substitutes subjects strictly prefer cooperation and the weak perfect substitutes subjects are indifferent between cooperation and defection, with all other subjects strictly preferring defection. Hence, our estimated preferences would predict between 6.25 and 22.4 percent cooperation, which roughly characterizes the findings.

These three examples do not, of course, prove that our results can explain all the findings of other studies, since most other studies differ in important ways from our own. However, the general ability of our results to characterize the findings elsewhere can, we believe, be taken as evidence that, overall, economic experiments are identifying a general degree of predictable and rational behavior, even when subjects are not money-maximizers.

9. JEALOUS PREFERENCES

Are preferences monotonic? There are several examples of violations of monotonicity in the experimental literature. Perhaps best known of these is the evidence of “disadvantageous counter-proposals” shown by Ochs and Roth (1989). In a multiple-round ultimatum bargain game, these authors observed subjects rejecting an offer in one round only to propose a division in the next round that provided less to both subjects than had the original offer been accepted. Another example is provided by Palfrey and Prisbrey (1996, 1997) who presented some subjects in a public goods game with a dominant strategy to contribute, which was not always taken. Is this behavior due to jealousy or spite, which implies nonmonotonic preferences, or is it a more complicated response to strategic concerns?

Table V shows the result of part 2 of session 5 which presented subjects with five upward sloping budgets. Subjects were constrained to the allocation of tokens listed in

TABLE V
CHOICES ON UPWARD SLOPING BUDGETS

	Budget				
	U1	U2	U3	U4	U5
<i>Definition of Budgets:</i>					
Self allocation in tokens	130	110	50	20	10
Other allocation in tokens	10	20	50	110	130
<i>Results:</i>					
Average valuation per token ^a	9.94	9.76	9.71	9.03	8.97
Standard deviation	0.3	0.9	1.2	2.1	2.6
Number of valuations < 10	1	3	2	8	7
Percent of subjects	2.9	8.8	5.9	23.5	20.6
Average valuation if < 10	8.0	7.3	5.0	6.4	5.0
Max	8	9	5	9	9
Min	8	5	5	2	0

^aSubjects choose to value all tokens from 0 to 10 cents each.

the first two rows of the table and could choose how much each token would be worth, from 0 to 10 cents. Note that of these five upward sloping budgets, two are advantageous (U1 and U2) and two are disadvantageous (U4 and U5). If preferences exhibit jealousy, then subjects could shrink the token values on U4 and U5 to gain less (absolute) inequality. It is possible, of course that subjects could even shrink token values on the advantageous budgets, U1 and U2. In this case, we might conjecture that choices illustrate humility.

Jealous preferences would mean that, as we move from left to right in the table, that is, more to less advantageous, the average valuation of tokens should get smaller. Indeed it does this, going from 9.94 cents to 8.97 cents. Overall, however, 88 percent of all choices are at the maximum. The nonmonotonicity is due to 8 subjects (23 percent). Given that subjects do shrink token values, the amount of shrinkage is somewhat severe, averaging 6.4 cents on U4 and 5 cents on U5. As expected, most of the nonmonotonic choices—71 percent—occur on the two disadvantageous budgets. Perhaps surprisingly, only one of the 34 subjects ever shrank the token value all the way to zero.

Looking at U1 and U2, we see distaste for inequality does not extend to advantageous inequality. U1 and U2 were shrunk at a quarter of the rate U4 and U5 were shrunk, and four of the five times these two budgets were shrunk the valuation was 8 or 9.

If preferences are nonmonotonic, but still convex, we can apply modified notions of revealed preference to the choices. Doing so, we find that four of the subjects making nonmonotonic choices do so in a way that is consistent with convexity, and four do not. Figure 5 gives an example of each type, where choices on upward sloping budgets are marked with circles. Subject 218 shows preferences that are convex and that dislike both extremes of inequality. The nonmonotonicity of subject 219 cannot be rationalized since the choice of *A* on the upward sloping budget cannot be reconciled with the choice *B* on the downward sloping budget.

10. CONCLUSION

Are altruistic choices consistent with the axioms of revealed preference such that a quasi-concave utility function could have generated the behavior? We find that it is indeed possible to capture altruistic choices with quasi-concave utility functions for individuals—altruism is rational. This is important for theories of fairness and altruism in experiments that are looking for a preference-based approach to explain the data.

What light can our findings shed on efforts to suggest utility functions for fairness and altruism? One essential observation from our study is that individuals are heterogeneous. There is clearly not one notion of fairness or inequality-aversion that all people follow—preferences range from Utilitarian to Rawlsian to perfectly selfish. Accounting for this difference will be a necessary part of understanding choices. A second critical observation is that fairness must be addressed and analyzed on an individual level. Because of the individual heterogeneity, a model that predicts well in the aggregate may not help us understand the behavior of individual actors. Capturing the variety of choices among individuals and then aggregating their behavior will lead to better understanding of both individuals and markets when altruism matters. Third, we found that a significant minority of subjects behave jealously—while maintaining convexity of preferences, they violate monotonicity. Fourth, our efforts to apply our results beyond simple dictator games suggests that many things other than the final allocation of money are likely to matter to subjects. Theories may need to include some variables from the game and the context in

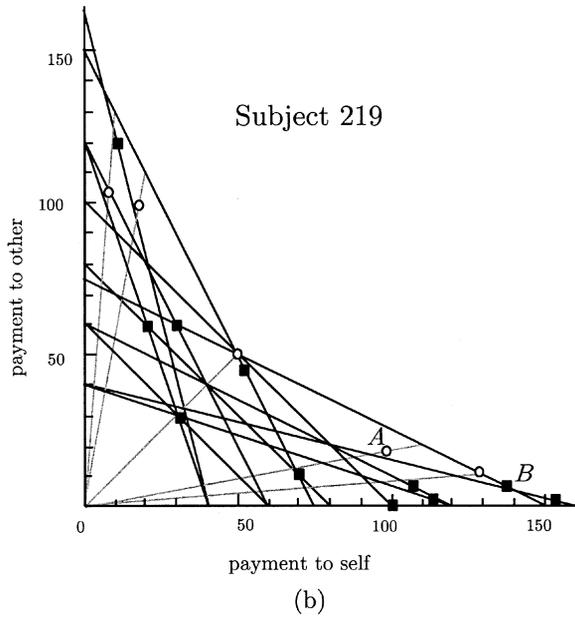
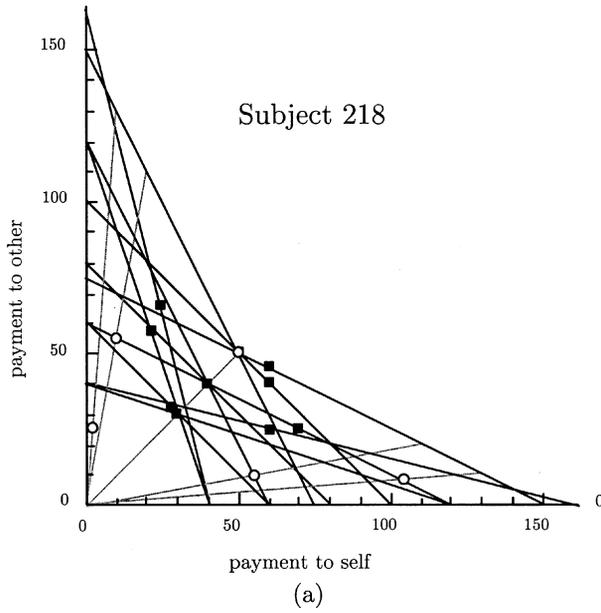


FIGURE 5.— Examples of nonmonotonic preferences. (a) Convex preferences. (b) Not rationalizable.

which the game is played if we are to understand the subtle influences on moral behavior like altruism.

Dept. of Economics, University of Wisconsin—Madison, 1180 Observatory Dr., Madison, WI 53706, U.S.A.; andreoni@facstaff.wisc.edu; www.ssc.wisc.edu/~andreoni

and

Dept. of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A., and Santa Fe Institute; miller@santafe.edu; zia.hss.cmu.edu

Manuscript received January, 1999; final revision received February, 2001.

REFERENCES

- AFRIAT, S. (1967): "The Construction of a Utility Function From Expenditure Data," *International Economic Review*, 8, 67–77.
- (1972): "Efficiency Estimates of Productions Functions," *International Economic Review*, 13, 568–598.
- ANDREONI, JAMES (1988): "Why Free Ride? Strategies and Learning in Public Goods Experiments," *Journal of Public Economics*, 37, 291–304.
- (1995a): "Warm-Glow versus Cold-Prickle: The Effect of Positive and Negative Framing on Cooperation in Experiments," *Quarterly Journal of Economics*, 109, 1–21.
- (1995b): "Cooperation in Public Goods Experiments: Kindness or Confusion?" *American Economic Review*, 85, 891–904.
- ANDREONI, JAMES, AND JOHN H. MILLER (1993): "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *Economic Journal*, 103, 570–585.
- BOHNET, IRIS, AND BRUNO S. FREY (1999a): "Social Distance and Other-regarding Behavior in Dictator Games: Comment," *American Economic Review*, 89, 335–339.
- (1999b): "The Sound of Silence in Prisoner's Dilemma and Dictator Games," *Journal of Economic Behavior and Organization*, 38, 43–57.
- BRONARS, STEPHEN G. (1987): "The Power of Nonparametric Tests of Preference Maximization," *Econometrica*, 55, 693–698.
- CASON, TIMOTHY N., AND VAI-LAM MUI (1997): "A Laboratory Study of Group Polarisation in the Team Dictator Game," *Economic Journal*, 107, 1465–1483.
- COX, JAMES C. (1997): "On Testing the Utility Hypothesis," *The Economic Journal*, 107, 1054–1078.
- CROSON, RACHEL T. A. (1996): "Partners and Strangers Revisited," *Economics Letters*, 53, 25–32.
- ECKEL, CATHERINE C., AND PHILIP J. GROSSMAN (1996): "Altruism in Anonymous Dictator Games," *Games and Economic Behavior*, 16, 181–191.
- FISHER, JOSEPH, R. MARK ISAAC, JEFFREY W. SCHATZBERG, AND JAMES M. WALKER (1995): "Heterogeneous Demand for Public Goods: Behavior in the Voluntary Contributions Mechanism," *Public Choice*, 85, 249–266.
- FORSYTHE, ROBERT, JOEL HOROWITZ, N. S. SAVIN, AND MARTIN SEFTON (1994): "Fairness in Simple Bargaining Games," *Games and Economic Behavior*, 6, 347–369.
- HARBAUGH, WILLIAM T., KATE KRAUSE, AND TIM BERRY (2001): "GARP for Kids: On the Development of Rational Choice Behavior," *American Economic Review*, 91, 1539–1545.
- HAUTHAKKER, HENRICK (1950): "Revealed Preference and the Utility Function," *Economica*, 17, 159–174.
- HOFFMAN, ELIZABETH, KEVIN MCCABE, AND VERNON SMITH (1996): "Social Distance and Other-Regarding Behavior in Dictator Games," *American Economic Review*, 86, 653–660.
- ISAAC, R. MARK, AND JAMES M. WALKER (1988): "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," *Quarterly Journal of Economics*, 53, 179–200.
- ISAAC, R. MARK, JAMES M. WALKER, AND ARLINGTON W. WILLIAMS (1994): "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups," *Journal of Public Economics*, 54, 1–36.

- LAURY, SUSAN K., JAMES M. WALKER, AND ARLINGTON W. WILLIAMS (1995): "Anonymity and the Voluntary Provision of Public Goods," *Journal of Economic Behavior and Organization*, 27, 365–380.
- MATTEI, AURELIO (2000): "Full-Scale Real Tests of Consumer Behavior Using Experimental Data," *Journal of Economic Behavior and Organization*, 43, 487–497.
- OCHS, JACK, AND ALVIN E. ROTH (1989): "An Experimental Study of Sequential Bargaining," *American Economic Review*, 79, 355–384.
- PALFREY, THOMAS R., AND JEFFREY E. PRISBREY (1996): "Altruism, Reputation and Noise in Linear Public Goods Experiments," *Journal of Public Economics*, 61, 409–422.
- (1997): "Anomalous Behavior in Public Goods Experiments: How Much and Why?" *American Economic Review*, 87, 829–846.
- SAMUELSON, PAUL A. (1938): "A Note on the Pure Theory of Consumer Behavior," *Economica*, 1, 61–71.
- SIPPEL, REINHARD (1997): "An Experiment on the Pure Theory of Consumer's Behavior," *The Economic Journal*, 107, 1431–1444.
- SONNEMANS, JOEP, ARTHUR SCHRAM, AND THEO OFFERMAN (1999): "Strategic Behavior in Public Good Games: When Partners Drift Apart," *Economics Letters*, 62, 35–41.
- VARIAN, HAL R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50, 945–972.
- (1991): "Goodness of Fit for Revealed Preference Tests," University of Michigan CREST Working Paper Number 13.
- (1993): *Microeconomic Analysis*, Third Edition. New York: Norton.
- WEIMANN, JOACHIM (1994): "Individual Behavior in a Free Riding Experiment," *Journal of Public Economics*, 54, 185–200.