Bounding the Impact of Market Experience on Rationality: Evidence from a Field Experiment with Imperfect Compliance*

John A. List
University of Maryland and NBER

Daniel L. Millimet
Southern Methodist University

3 April 2005

Abstract

While laboratory experiments documenting some level of irrational behavior are now commonplace, explorations into whether such irrationalities exist in the field are rare. Equally as scarce are studies that explore the influence of market experience on the level and evolution of irrationality. Using field data gathered from more than 380 subjects of age 6-18, we investigate these issues using Generalized Axiom of Revealed Preference experiments. To circumvent the endogeneity of market experience, we exogenously induce such experience through the design of a field experiment. Compliance with the experiment was not perfect, however. We are, nevertheless, able to bound the average treatment effect using the sharp bounds derived in Balke and Pearl [*Journal of the American Economic Association*, 1997, 92, 1171-1776]. Empirical results indicate that deviations from rational behavior exist in the field, but that market experience is a significant contributor to the development of rational choice.

JEL: C14, C93
Key words: rationality, market learning, field experiment, imperfect compliance, treatment effects, nonparametric bounds, instrumental variables, intent-to-treat

---

*Corresponding author: Daniel Millimet, Department of Economics, Box 0496, Dallas, TX 75275-0496, USA; Email: millimet@mail.smu.edu; Telephone: 214.768.3269; Fax: 214.768.1821. Thanks to Mike Price for research assistance. George Akerlof, James Andreoni, William Harbaugh, Liesl Koch, Marc Nerlove, and Vernon Smith provided helpful insights during this research project. The authors are also thankful from constructive comments from Patrick Bajari, James Cox, Dan Houser, Charles Manski, and other participants at the Princeton Experimental Economics Conference, March 31 – April 1, 2005.

I. Introduction

The assumption of individual rationality is one of the fundamental pillars, if not *the* pillar, upon which the majority of economic models are based. The rationality of individual decision-making has long been under a cloud of suspicion, however. For instance, the historical work of William Stanley Jevons, Irving Fisher, Alfred Marshall, and A.C. Pigou openly questioned the rationality of individual intertemporal consumption decisions, and Baudin (1954, p. 493) exclaims matter-of-factly that "it is known that the consumer is an irrational being." More recently, empirical evidence obtained from controlled laboratory experiments combined with Varian's (1982) theory of revealed preference highlights the level of individual irrationality (see, e.g., Sippel 1997; Harbaugh et al. 2001; Andreoni and Miller 2002).[1] Such studies suggest that between 10 and 75 percent of experimental subjects violate the Generalized Axiom of Revealed Preference (GARP).[2]

Despite the importance of these laboratory studies, whether and to what extent agents in the field exhibit similar irrationalities is largely unknown. Furthermore, evidence on what accounts for the variation in the level of rationality across individuals as well as if, and how, individuals evolve into more rational beings – a concept we refer to as *endogenous rationality* – is minimal at best. Brief discussions in the extant literature focus on education and market experience; specifically, Peart (2000, p. 188) brings to the forefront the arguments of the prominent early neoclassical economists in support of educating the impoverished in order to reduce the "mistakes" of the "imperfect decision maker." In a related literature, Koopmans

[1] Throughout the paper, we use the term "irrationality" to mean, more precisely, deviations from the Generalized Axiom of Revealed Preference.

[2] Sippel (1997) uses students of law or economics and ten budget sets for eight goods. Mattei (2000) uses college students and other adults, along with 20 budget sets for eight goods. Harbaugh et al. (2001) use 7- and 11-year-old participants and eleven budget sets for two goods. Andreoni and Miller (2002) use college economics students and eleven budget sets defined in a modified dictator game where subjects had to decide how much money to keep and how much to give to charity. The use of experiments to test individual behavior actually has a much longer history (see, e.g., May (1954), MacCrimmon and Toda (1969), Battalio et al. (1973)).

(1964) and, more recently, Bowles (1998) emphasize the role of market experience and economic institutions in the evolution of individual *preferences*, as opposed to rationality. Yet because the analyses of Koopmans, Bowles, and others concerned with endogenous preferences are predicated upon individuals making rational decisions at any point in time, the impact of market experience and economic institutions on endogenous preferences may (in part) reflect their influence on individual rationality.

The goals of this paper are threefold: i) to explore the extent of irrationality in the field, ii) to examine the role of market experience on the *level* of individual rationality, and iii) to estimate the role of market experience on the *evolution* of individual rationality. In all cases, we follow the recent literature and measure rationality using a Generalized Axiom of Revealed Preference (GARP) experiment. To accomplish these goals, while allowing for the fact that preferences might evolve over time, we construct a panel data set across young individuals using a controlled field experiment. The two rounds of our experiment take place seven months apart and include children 6-18 years old at a shopping mall. Since each round of the experiment involves subjects making consumption choices under different budget constraints, the level of rationality exhibited by subjects during each round is identifiable, even if respondents' preferences change over time.

To assess the impact of market experience on the *level* and *evolution* of individual rationality, while recognizing that market experience itself is most likely endogenous, we randomly induce participation by some of the subjects into the market for sports memorabilia in between the two experimental rounds. This approach represents a particularly demanding test of the market's role on rationality since it represents a test of how experience in one well-defined market affects rational choice behavior in a *separate*, quite distinct venue (rather than a test of how experience in a particular market induces certain heuristics or "rules of thumb" that can be applied in future transactions in the *same* market over similar tasks).

Although seemingly well conceived, as in many non-laboratory studies designed to elicit causal effects of a particular treatment via randomization, compliance by the experimental subjects was imperfect. Consequently, we utilize the technique of Balke and Pearl (1997) to provide sharp bounds of the average treatment effect under minimal assumptions. In addition, we utilize an instrumental variable (IV) technique based on the "intent-to-treat" to provide estimates of the average impact of market experience on both the *level* of and *change* in rationality displayed by the sub-population of subject compliers.

Empirical results provide robust evidence that market experience facilitates the development of rational behavior. Specifically, we find that exogenous inducement into the sports memorabilia market, if applied universally to the entire population, would decrease the probability of irrational behavior (i.e., non-zero GARP violations) by -1.6% to 29.0%. The IV "intent-to-treat" estimate is 19.7%. Although the nonparametric bound for the level of GARP violations includes zero, the bound for the evolution of rational choice does not. Specifically, we find that exogenous inducement into the sports memorabilia market, if applied universally to the entire population, would increase the probability of becoming more rational (i.e., reduce GARP violations) by 6.4% to 36.9%. The IV "intent-to-treat" estimate is 23.1%.

The remainder of our study is organized as follow. Section II describes the experimental protocol. Section III analyzes the results. Section IV concludes.

II. Experimental Design

The treatments were conducted at a shopping mall in a large southern city from November 2000 to June 2001 during times when a sports memorabilia show was in progress. In the first round of the experiment undertaken in November of 2000, the monitor approached *young* individuals in and near the sports memorabilia marketplace and inquired about their interest in participating in an experiment that would take about ten minutes. The decision to solicit youths as experimental subjects stemmed from our desire to examine the role of the market in the development of rational choice behavior.

If the individual agreed to be an experimental participant, the monitor began the four steps of the experiment: first, the subject began by filling out a brief survey in which information on age, gender, years of sports memorabilia market experience, buying, selling, and trading intensity in the sports memorabilia market, and the number of monthly visits to the mall were obtained. After completing the survey, Step 2 began: the subject was physically given the experimental sheets and instructions for the GARP experiment. Our GARP treatments closely follow Harbaugh et al. (2001), as we present our subjects with 11 different choice sets (over boxes of juice and bags of chips) on 11 separate sheets of paper and inform each subject to choose their most preferred bundle on each sheet, after which the monitor will choose one sheet to execute.[3]

Figure 1 graphically depicts the 11 choice sets. In our design, a GARP violation occurs when a bundle *x* is chosen when a bundle *y* is available, where bundle *y* has at least as much of all goods and strictly more of at least one good than a third bundle *z*, and *z* has been directly or indirectly revealed preferred to *x*. A bundle *z* is directly revealed preferred to another bundle *x* if (i) *z* is chosen when *x* is available, or (ii) *z* is chosen when another bundle containing at least as much of all goods as in *x* and strictly more of at least one good is available. Bundle *z* is

[3] We should highlight that while we believe GARP is the most fundamental definition of rationality, one could use several other measures of rationality. In this spirit, it is possible that subjects who have few (many) violations in our setting have many (few) violations of rationality in other less complicated settings.

indirectly revealed preferred to bundle *x* if a string of directly preferred relations suggests that *z* is preferred to *x*. If an individual's choices do not violate GARP, they are consonant with the individual maximizing a continuous, concave, strongly monotonic utility function.

Several examples were carried out to ensure that the subject understood the details of the experiment. No time limit was imposed. In Step 3, the monitor informed the subject which sheet was to be actually executed and the subject received his or her chosen bundle from that sheet. Step 4 closed the experiment and included "parting gifts" for subjects in certain treatments (explained more fully below).

To assess the causal effect of market experience on rational choice, for the remainder of the paper we focus on only those subjects who were simply at the mall for enjoyment and not to participate in the sports memorabilia market.[4] To proceed, we delineated the subjects without previous sports memorabilia market experience into two groups: *GIFT* and *NOGIFT*. The delineation was changed at the top of each hour, so subjects' treatment type was determined exogenously based on the time they visited the mall.

For the *GIFT* subjects, we provided a "parting gift" of approximately $25 worth of sports cards and memorabilia in Step 4 of the experiment. In each subject's gift bundle, we included several items designed to engage the subject in the marketplace. The monitor informed subjects in this treatment group that the gifts were theirs to keep, and they could sell or trade the gifts in the marketplace or take them home. The monitor stressed that dealers at the show were interested in the goods and that the goods had a book value of approximately $25. Furthermore, in an effort to ensure that the local dealers would have an interest in buying and trading with these subjects, prior to the show the monitor discussed various items with the dealers attending

[4] Thus, these subjects answered "none" to all the questions concerning previous experience in the sports memorabilia market in November 2000. In a companion study (List and Millimet, 2004), we study the behavior of these other types of subjects.

the show to determine the appropriate composition of the gift packages. In practice, these agents engaged in buying, selling, and trading in the live marketplace. Treatment *NOGIFT* was identical to the *GIFT* treatment except that in Step 4 the monitor closed the experiment by thanking the subject for his or her participation (and no "parting gift" was given).

To complete the experiment, we returned to the same mall the following June and ran similar GARP treatments using the same subject pool, identical experimental procedures, but different goods (instead of chips and juice boxes we used packs of gum and candy bars). To recruit the same subjects, in May one of the authors personally telephoned and/or e-mailed the 219 subjects that participated in the November 2000 experiment. He was able to contact and obtain agreement to meet him at the June sportscard show from roughly two-thirds of the subjects. As a friendly reminder, within two weeks of the experiment he called/e-mailed the subjects that agreed to participate in the follow-up; despite this reminder only 73 subjects attended round two of the experiment.

Table 1 provides a summary of the experimental design and the number of subjects in each treatment. The *GIFT* treatment in November included 110 subjects and of those 110 subjects 42 returned for the second GARP experiment, which we denote treatment *GIFTII*. Of the returning participants, not all complied with the intentions of the experiment. In particular, of the 42 *GIFTII* subjects, two did not enter the sports memorabilia market to sell their parting gift; of the 31 *NOGIFTII* subjects, eight subjects voluntarily entered the sports memorabilia market between the two experimental rounds. Thus, we refer to the 40 market participants in *GIFTII* and the 23 non-participants in *NOGIFTII* as "compliers" and the remaining ten subjects as non-compliers.

III. Empirical Methodology

A. Sharp Bounds for the Average Treatment Effect

To assess the average causal effect of market participation on the *level* and *change* in rational choice, we begin by utilizing the bounding method of Balke and Pearl (1997). Let Z_i denote the randomized treatment assignment for subject *i*; D_i denotes the actual treatment received. Let Y_i denote the observed outcome of interest for subject *i*. *Z*, *D*, and *Y* are binary variables. For *Z* and *D*, one (zero) denotes market participants (non-participants). For *Y*, one (zero) denotes a "positive" ("negative") outcome. Specifically, when analyzing the level of GARP violations in round two, Y_i equals one (zero) if the subject incurs no (some) GARP violations. When analyzing the change in GARP violations between the two rounds, Y_i equals one (zero) if the number of GARP violations by the subject decreased (increased or remained unchanged).

To bound the average treatment effect, we maintain the following assumptions:

(A1) *Conditional Independence*: $Y \perp\!\!\!\perp Z \mid \{D, U\}$

(A2) *Marginal Independence*: $U \perp\!\!\!\perp Z$

where U_i represents observed and unobserved attributes of subject *i*. (A1) states that the randomized treatment assignment has no direct effect on *Y* conditional on the actual treatment received (and *U*); Angrist et al. (1996) refer to this as the "exclusion restriction" assumption. (A2) requires that *Z* be randomly assigned and that *U* not be affected by *Z*; hence, Angrist et al. (1996) refer to this as the assumption of "random assignment."

Under (A1) and (A2), which appear innocuous given the experimental design detailed in the previous section, one may decompose the joint distribution $P(y, d, z, u)$ in the following manner:

$$P(y, d, z, u) = P(y \mid d, u)\, P(d \mid z, u)\, P(z)\, P(u) \tag{1}$$

where lower case values denote specific values of Y, D, Z, and U (and U may be continuous despite the use of $P(\cdot)$ to denote its distribution).

The joint distribution in (1) is not directly observed given the presence of U, but the probabilities $P(y, d \mid z = 0)$ and $P(y, d \mid z = 1)$ are observable. Given this information, one can bound the following expressions of interest:

$$P(y \mid \breve{d}) = \sum_{u} P(y \mid d, u) P(u) \tag{2}$$

$$\begin{aligned} \tau &= P(y = 1 \mid \breve{d} = 1) - P(y = 1 \mid \breve{d} = 0) \\ &= \sum_{u} P(u) \left[P(y = 1 \mid d = 1, u) - P(y = 1 \mid d = 0, u) \right] \end{aligned} \tag{3}$$

where $\breve{d}$ indicates that the value of d has been determined by complete randomization (i.e., perfect compliance). Expression (2) represents the probability that Y takes on a specific value if treatment assignment, D, was completely determined by a randomized experiment. Expression (3) represents the average treatment effect, equal to the difference in the probability that Y takes on the "positive" outcome given random assignment to the treatment versus the control group.

Following Balke and Pearl (1997), we use the following notation to define the observed probabilities:

$p_{00.0} = P(y = 0, d = 0 \mid z = 0)$ $\quad$ $p_{00.1} = P(y = 0, d = 0 \mid z = 1)$

$p_{01.0} = P(y = 0, d = 1 \mid z = 0)$ $\quad$ $p_{01.1} = P(y = 0, d = 1 \mid z = 1)$

$p_{10.0} = P(y = 1, d = 0 \mid z = 0)$ $\quad$ $p_{10.1} = P(y = 1, d = 0 \mid z = 1)$

$p_{11.0} = P(y = 1, d = 1 \mid z = 0)$ $\quad$ $p_{11.1} = P(y = 1, d = 1 \mid z = 1)$

This setup yields the following bounds for (2):

$$P(y=1 \mid \breve{d}=0) \in \left[\max\begin{Bmatrix} p_{10.1} \\ p_{10.0} \\ p_{10.0}+p_{11.0}-p_{00.1}-p_{11.1} \\ p_{01.0}+p_{10.0}-p_{00.1}-p_{01.1} \end{Bmatrix}, \min\begin{Bmatrix} 1-p_{00.1} \\ 1-p_{00.0} \\ p_{01.0}+p_{10.0}+p_{10.1}+p_{11.1} \\ p_{10.0}+p_{11.0}+p_{01.1}+p_{10.1} \end{Bmatrix} \right] \quad (4)$$

$$P(y=1 \mid \breve{d}=1) \in \left[\max\begin{Bmatrix} p_{11.0} \\ p_{11.1} \\ -p_{00.0}-p_{01.0}+p_{00.1}+p_{11.1} \\ -p_{01.0}-p_{10.0}+p_{10.1}+p_{11.1} \end{Bmatrix}, \min\begin{Bmatrix} 1-p_{01.1} \\ 1-p_{01.0} \\ p_{00.0}+p_{11.0}+p_{10.1}+p_{11.1} \\ p_{10.0}+p_{11.0}+p_{00.1}+p_{11.1} \end{Bmatrix} \right] \quad (5)$$

The bounds for τ in (3) are then functions of the bounds given in (4) and (5). Specifically, the lower bound for τ is equal to the difference between the lower bound in (5) and the upper bound in (4); the upper bound for τ is given by the difference between the upper bound in (5) and the lower bound in (4).

The bounds on τ are superior to those provided in Robins (1989) and Manski (1990) in the event that monotonicity does not hold; if monotonicity does hold, the bounds are identical. Monotonicity requires that subjects do not consistently act contrary to their (random) treatment assignment (Angrist et al. 1996). Formally, $P(d = 1 \mid z = 1, u) \geq P(d = 1 \mid z = 0, u)$ for all u. Without monotonicity, the width of the bounds is no greater than the rate of noncompliance, $P(d = 1 \mid z = 0) + P(d = 0 \mid z = 1)$ and may collapse to a point estimate in certain cases. With the monotonicity assumption, the width of the bounds is exactly equal to the rate of noncompliance.

Lastly, Balke and Pearl (1997) note that the requirement that the lower bounds be smaller than their corresponding upper bounds implies certain testable restrictions. Without monotonicity being imposed, the following inequalities must hold:

$$p_{00.0}+p_{10.1} \leq 1$$

$$p_{01.0}+p_{11.1} \leq 1$$

$$p_{10.0}+p_{00.1} \leq 1$$

$$p_{11.0} + p_{01.1} \leq 1 \qquad (6)$$

Under monotonicity, the inequalities in (6) may be replaced with

$$p_{01.0} \leq p_{01.1}$$

$$p_{11.0} \leq p_{11.1}$$

$$p_{00.1} \leq p_{00.0}$$

$$p_{10.1} \leq p_{10.0} \qquad (7)$$

Depending on whether one assumes monotonicity to hold or not, violation of (6) or (7) suggests that (A1) and/or (A2) does not hold.

B. Point Estimates of the Average Treatment Effect

To complement the bounds, we also report point estimates derived via two methods. The first is the familiar, but potentially misleading, "intent-to-treat" estimator, denoted herein by $\hat{\tau}_{ITT}$. Using the previous notation, $\hat{\tau}_{ITT}$ is calculated as

$$\hat{\tau}_{ITT} = P(y = 1 \mid z = 1) - P(y = 1 \mid z = 0) \qquad (8)$$

Equation (8) represents the mean difference in outcomes across subjects delineated by treatment assignment, irrespective of the treatment actually received.

Another estimator, proposed in Angrist et al. (1996), uses Z as an instrumental variable (IV) for D. The IV (Wald) estimator, denoted by $\hat{\tau}_{IV}$, is given by

$$\hat{\tau}_{IV} = \frac{P(y = 1 \mid z = 1) - P(y = 1 \mid z = 0)}{P(d = 1 \mid z = 1) - P(d = 1 \mid z = 0)} = \frac{\hat{\tau}_{ITT}}{\hat{\tau}_{TS}} \qquad (9)$$

which is equal to the ratio between the "intent-to-treat" estimator and the average causal effect of treatment assignment, Z, on actual treatment status, denoted by $\hat{\tau}_{TS}$. The IV estimator requires the following assumptions in addition to (A1) and (A2)

(A3) *Stable Unit Treatment Value Assumption*: $P(d_i \mid z_i, z_j) = P(d_i \mid z_i)$ and $P(y_i \mid d_i, d_j, z_i, z_j) = P(y_i \mid d_i, z_i)\ \forall j \neq i$

(A4) *Nonzero Average Causal Effect of Z on D*: $P(d \mid z=1) - P(d \mid z=0) \neq 0$

(A5) *Monotonicity*: $P(d = 1 \mid z = 1) \geq P(d = 1 \mid z = 0)$

(A3) requires that the potential outcomes of each subject be independent of the treatment assignment and actual treatment status of all other subjects. (A4) requires that actual treatment status be related to the original treatment assignment. (A5) is defined above.

Finally, before continuing, it is important to note that in a heterogeneous treatment effect framework, the bounding and IV estimators do not estimate the same parameter. Specifically, if the treatment effect varies across subjects, then the IV estimator reflects the average treatment effect for the sub-population of compliers, whereas the bounds reflect the average treatment effect for the entire population. On the other hand, if the treatment effect is constant across subjects, then the treatment effect in the sub-population of compliers is identical to the population treatment effect, and the bounding and IV estimators measure the same parameter.

IV. Results

The frequency counts used to construct the empirical conditional probabilities used to form the Balke and Pearl (1997) bounds are given in Table 2; Table 3 reports the empirical conditional probabilities. In each table the results in Panel I refer to the change in GARP violations between rounds, and the results in Panel II refer to the level of GARP violations in round two. Figure 2 plots the distribution of GARP violations for subjects in round two.

In terms of the average causal effect of market experience on the learning of rational behavior, the lower (upper) bound is 0.064 (0.369). Thus, the bounds exclude zero, and indicate that if market experience were uniformly obtained by the entire population, the probability that individuals would make more rational choices would increase by no less than 6.4% and no more than 36.9%.[5] This result is compelling, offering robust evidence in favor of one mechanism by which individuals learn to exhibit more rational behavior.

In terms of the level of GARP violations, the results are less conclusive. Now, the lower (upper) bound is -0.016 (0.290), which includes zero. The bounds imply that if market experience were uniformly obtained by the entire population, the probability that individuals would cease to exhibit any GARP violations would increase by no more than 29.0% and would decline by no more than 1.6%.

Before assessing the point estimates, we examine the various testable implications detailed in the previous section. First, we note that the assumption of monotonicity appears to hold in the data; thus, the improvement in the Balke and Pearl (1997) bounds over the bounds in Robins (1989) and Manski (1990) is inconsequential in the present analysis. Given that monotonicity holds, one will also notice that the width of the bounds reported in Table 4 is equal to the rate of noncompliance, 0.306. Second, we note that all the inequalities listed in (6) and (7) are met in the data.

Turning to the point estimates, the "intent-to-treat" estimator, $\hat{\tau}_{ITT}$, is 0.161 (standard error = 0.118) for the change in GARP violations and 0.137 (standard error = 0.119) for the level of GARP violations. Both clearly lie within the Balke and Pearl (1997) bounds, but are a potentially very misleading. In terms of learning rational behavior, the "intent-to-treat" estimator may overstate (understate) the impact of market experience by roughly 9.7% (20.8%); for the

[5] 90% confidence intervals for the bounds are also reported using the method in Lechner (1999). The method entails bootstrapping the bounds and reporting the 5^{th} (95^{th}) percentile of the lower (upper) bound. Note, however, that the proper interpretation is that there is approximately a 10% probability that the true *interval* for τ is not contained within the bootstrap bounds; the probability that the true *value* of τ lies outside the interval defined by the bootstrap values is necessarily much lower than 10%. The bootstrap values are obtained from 10,000 repetitions.

level of rational behavior, it may overstate (understate) the impact of market experience by roughly 15.3% (15.3%).

The IV estimator, $\hat{\tau}_{IV}$, which is simply the "intent-to-treat" estimator scaled by the average causal effect of treatment assignment on actual treatment status, $\hat{\tau}_{TS}$, is 0.231 (standard error = 0.164) for learning and 0.197 (standard error = 0.168) for the level. As with the "intent-to-treat" estimator, the IV estimator may overstate (understate) the (population) impact of market experience by roughly 16.7% (13.8%); for the level of rational behavior, it may overstate (understate) the (population) impact of market experience by roughly 21.3% (9.3%). In the present case, however, the additional assumptions required of the IV estimator, namely (A4) – (A5), hold in the data, and there is little reason to suspect (A3) is violated given the experimental design. Thus, one might be willing to accept the results of IV estimator, indicating an average treatment effect (for the sub-population of compliers) of market experience that lie in the upper portions of the bounds (for the average treatment effect).

V. Conclusion

The fact that economic agents in the laboratory do not consistently exhibit behavior consonant with rational choice is well-documented. Yet, our understanding of the variation in rational behavior across individuals in the field, as well as our knowledge of the mechanisms by which agents become more rational over time, is minimal. This study assesses the role played by the market in endogenous development of rational behavior. To that end, we present evidence from a set of experimental treatments that exogenously induce market participation by individuals, finding that the market is more powerful than most surmise: using a straightforward test for rationality relying on the theory of revealed preference, we find robust evidence that market experience is one mechanism by which individuals learn to become more rational. This result, based on the bounding technique of Balke and Pearl (1997) which relies on minimal assumptions, is surprising in that our experimental design constitutes a particularly challenging test of learning, as our analysis quantifies the average causal impact of participation in one market on the development of rational choice behavior in a *separate*, quite distinct, “market.”

References

Andreoni, J. and J. Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 70, 737-53.

Angrist, J.D., G.W. Imbens, and D.B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444-472.

Balke, A. and J. Pearl. 1993. "Nonparametric Bounds on Causal Effects From Partial Compliance Data," Technical Report R-199, UCLA, Department of Computer Science.

Balke, A. and J. Pearl. 1994. "Counterfactual Probabilities: Computational Methods, Bounds and Applications," in *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, San Francisco: Morgan Kauffman.

Balke, A. and J. Pearl. 1997. "Bounds on Treatment Effects From Studies with Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171-1176.

Battalio, R.C., J.H. Kagel, R.C. Winkler, E.B. Fisher, R.L. Basmann, and L. Krasner. 1973. "A Test of Consumer Demand Theory Using Observations of Individual Consumer Purchases," *Western Economic Journal*, 11, 421-8.

Baudin, L. 1954. "Irrationality in Economics," *Quarterly Journal of Economics*, 68, 487-502.

Bowles, S. 1998. "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions," *Journal of Economic Literature*, 36, 75-111.

Harbaugh, W.T., K. Krause, and T.R. Berry. 2001. "GARP for Kids: On the Development of Rational Choice Behavior," *American Economic Review*, 91, 1539-45.

Koopmans, T.C. 1964. "On Flexibility of Future Preferences," in S. Maynard and B. Glenn (eds.) *Human Judgments and Optimality*, New York: Wiley.

Lechner, M. 1999. "Nonparametric Bounds on Employment and Income Effects of Continuous Vocational Training in East Germany," *Econometrics Journal*, 2, 1-28.

List, J.A., and D.L. Millimet. 2004. "The Market: Catalyst for Rationality and Filter of Irrationality," working paper.

MacCrimmon, K.R. and M. Toda. 1969. "The Experimental Determination of Indifference Curves," *Review of Economic Studies*, 36, 433-51.

Mattei, A. 2000. "Full Scale Real Tests of Consumer Behavior Using Experimental Data," *Journal of Economic Behavior and Organization*, 43, 487-97.

May, K.O. 1954. "Intransitivity, Utility, and the Aggregation of Preference Patterns," *Econometrica*, 22, 1-13.

Peart, S.J. 2000. "Irrationality and Intertemporal Choice in Early Neoclassical Thought," *Canadian Journal of Economics*, 33, 175-189.

Robins, J.M. 1989. "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in L. Sechrest, H. Freeman, and A. Mulley (eds.) *Health Service Research Methodology: A Focus on AIDS*, Washington D.C.: U.S. Public Health Service.

Sippel, R. 1997. "An Experiment on the Pure Theory of Consumers' Behavior," *Economic Journal*, 107, 1431-44.

Slonim, R.L. 1999. "Learning Rules of Thumb or Learning More Rational Rules," *Journal of Economic Behavior and Organization*, 38, 217-36.

Varian, H. 1982. "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50, 945-72.

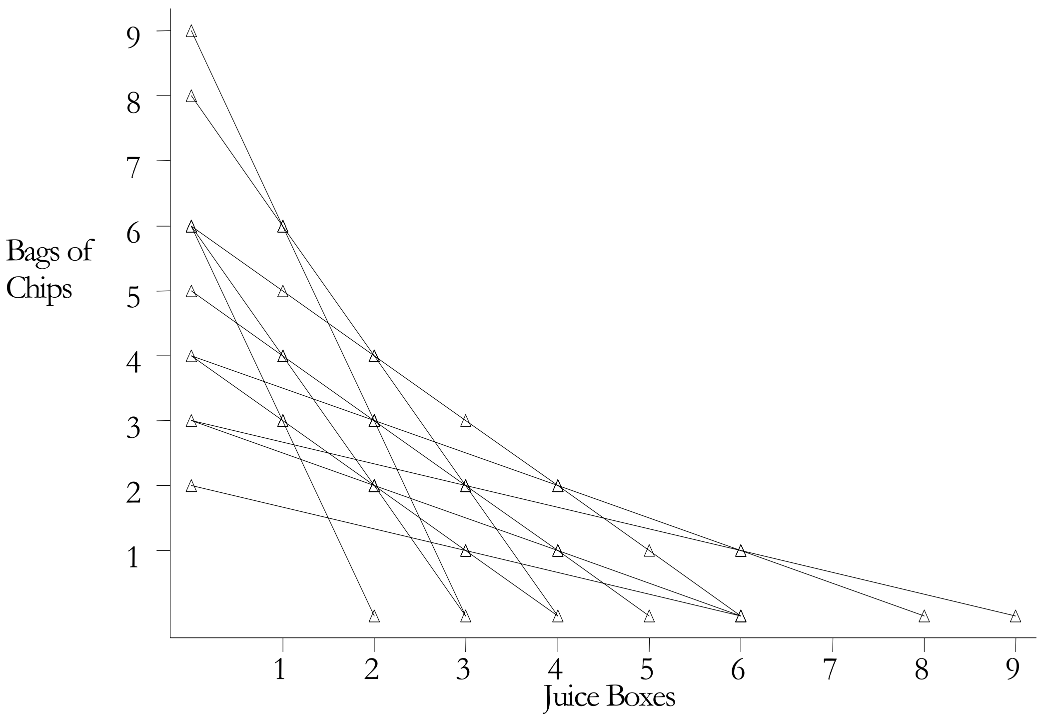

Figure 1. Choice Sets

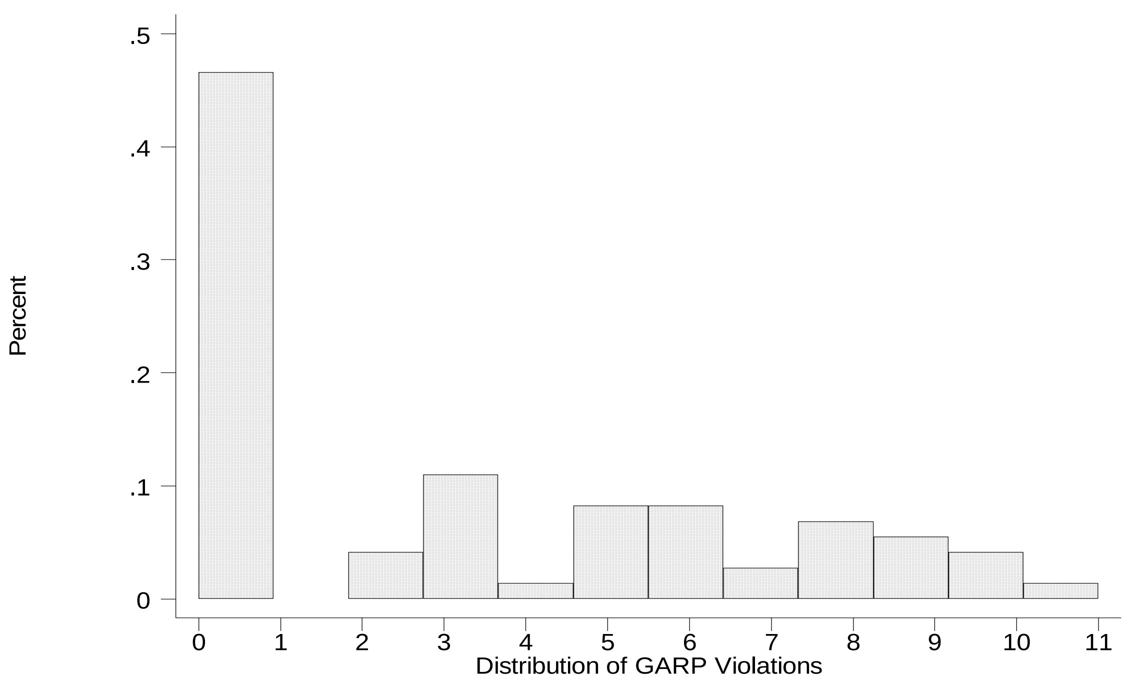

Figure 2. GARP Violations

Table 1. Individual Choice (GARP) Experimental Design

	November 2000	**June 2001**
Subject with free gift bag	*GIFT* n = 110	*GIFTII* n = 42
Subject without free gift bag	*NOGIFT* n = 109	*NOGIFTII* n = 31

Table 2. Count of Subjects According to Treatment Assignment, Treatment Received, and Outcome.

	$z = 0$		$z = 1$	
	$y = 0$	$y = 1$	$y = 0$	$y = 1$
I. Outcome: Change in GARP Violations				
$d = 0$	16	7	2	0
$d = 1$	3	5	17	23
II. Outcome: Level of GARP Violations				
$d = 0$	15	8	1	1
$d = 1$	4	4	19	21

Notes: z refers to treatment assignment (0 = *NOGIFT*; 1 = *GIFT*); d refers to treatment received (0 = *NOGIFT*; 1 = *GIFT*); y refers to the outcome (0 = negative outcome; 1 = positive outcome). In Panel I, $y = 1$ if GARP violations decreased from the first to second round, 0 otherwise. In Panel II, $y = 1$ if zero GARP violations were reported in the second round, 0 otherwise.

Table 3. Empirical Conditional Probability Distribution $P(y,d|z)$.

	$z = 0$		$z = 1$	
	$y = 0$	$y = 1$	$y = 0$	$y = 1$
I. Outcome: Change in GARP Violations				
$d = 0$	0.516	0.226	0.048	0.000
$d = 1$	0.097	0.161	0.405	0.548
II. Outcome: Level of GARP Violations				
$d = 0$	0.484	0.258	0.024	0.024
$d = 1$	0.129	0.129	0.452	0.500

Notes: Calculations based on data provided in Table 2.

Table 4. Bounds and Point Estimates for the Average Treatment Effect.

	Change in GARP Violations			Level of GARP Violations		
	Lower	**Upper**	**Width**	**Lower**	**Upper**	**Width**
Bounds	0.064	0.369	0.306	-0.016	0.290	0.306
	[-0.130]	[0.544]		[-0.210]	[0.467]	
τ_{ITT}	0.161	0.161	0.000	0.137	0.137	0.000
	(0.118)	(0.118)		(0.119)	(0.119)	
τ_{TS}	0.694	0.694	0.000	0.694	0.694	0.000
	(0.079)	(0.079)		(0.079)	(0.079)	
τ_{IV}	0.231	0.231	0.000	0.197	0.197	0.000
	(0.164)	(0.164)		(0.168)	(0.168)	

Notes: Standard errors in parentheses. 5th (95th) percentile of the bootstrap distribution for the lower (upper) bound, based on 10,000 repetitions, reported in brackets.