TESTING CONSUMER THEORY IN THE FIELD: PRIVATE CONSUMPTION VERSUS CHARITABLE GOODS*

STEFFEN HUCK† IMRAN RASUL‡

UNIVERSITY COLLEGE LONDON AND ELSE

OCTOBER 2008

Abstract

We present evidence from a natural field experiment designed to shed light on whether individual behavior is consistent with a neoclassical model of utility maximization subject to budget constraints. We do this through the lens of a field experiment on charitable giving. In conjunction with the Bavarian State Opera House, we mailed 25,000 regular opera attendees a letter describing a charitable fundraising project organized by the opera house. Recipients were randomly assigned into one of five treatments designed to provide ten tests of revealed preference theory. We find that the behavior of at least 80% of individuals, on both the extensive and intensive margins, can be rationalized within a standard neoclassical choice model in which individuals have preferences defined over own consumption and their contribution towards the charitable good, and their preferences satisfy the axioms of revealed preference. The analysis highlights that in a real world environment in which individuals make simple decisions they are familiar with, standard microeconomic theory works well in explaining observed individual choices.

Keywords: consumer theory; natural field experiment; revealed preference.

JEL Classification: C93, D01, D12, D64.

---

*We thank Sami Berlinski, Stéphane Bonhomme, Guillermo Caruana, Syngjoo Choi, Heike Harmgart, Dean Karlan, Enrico Moretti, Sendhil Mullainathan, Adam Rosen, Georg Weizsäcker, and seminar participants at Autonoma, CEMFI, LSE, and the LEaF 2007 Conference at UCL for useful comments. We gratefully acknowledge financial support from the ESRC. All errors remain our own.

†Department of Economics and the ESRC Centre for Economic Learning and Social Evolution, University College London, Drayton House, 30 Gordon Street, London WC1E 6BT, United Kingdom. Tel: +44-207-679-5895; Fax: +44-207-916-2775; E-mail: s.huck@ucl.ac.uk.

‡Department of Economics and the ESRC Centre for Economic Learning and Social Evolution, University College London, Drayton House, 30 Gordon Street, London WC1E 6BT, United Kingdom. Tel: +44-207-679-5853; Fax: +44-207-916-2775; E-mail: i.rasul@ucl.ac.uk.

1 Introduction

Neoclassical consumer theory provides a rich set of testable implications for how consumer demand responds to changes in relative prices and income. This paper presents evidence from the first large-scale natural field experiment designed to shed light on whether individual behavior is consistent with the predictions of revealed preference theory within a standard model of utility maximization subject to budget constraints [Samuelson 1947, Houthakker 1950, Afriat 1967]. We do this through the lens of a natural field experiment on charitable giving. In conjunction with the Bavarian State Opera House in Munich, in June 2006 we mailed 25,000 regular opera attendees a letter describing a charitable fundraising project organized by the opera house. Recipients were randomly assigned to one of five treatments designed to test revealed preference theory.[1]

By focusing our analysis on the choice between a charitable good and all other private consumption, we are able to vary the budget set any given individual faces in straightforward and natural way, holding all other prices constant. We do so by simply offering various matching schemes that affect how donations given for the charitable good translate into donations actually received by the project. Specifically, we induce—(i) large changes in the relative price of the charitable good through various rates at which donations are matched; (ii) pure income transfers to individuals through a matching scheme that guarantees *any* positive donation is matched by some fixed amount; (iii) non-convex budget set in which the donation is only matched if it is above a fixed threshold.

Hence our treatments experimentally induce changes in the relative price of the charitable good relative to own consumption, holding constant all other prices, asset holdings, and labor supplies, as well as inducing small changes in income, again holding everything else equal. These are precisely the types of change in budget set that the axioms of revealed preference predict behavioral responses to. The fact that the induced budget sets intersect each other, then opens up the possibility to directly test the predictions of revealed preference theory.

Our between-subject experimental research design avoids some of the empirical concerns that would arise with any within-subject design. For example, within-subject designs inevitably require the same individual to be presented with different budget sets at different moments in time. This raises the concern that there are natural changes over time in incomes, relative prices, asset holdings or labor supplies, that confound any inference that can be made on whether individual preferences satisfy the axioms of revealed preference. Such concerns do not arise in our between-subject design.

Our approach borrows from the program evaluation literature [Heckman *et al* 1999]. Under the twin identifying assumptions of random assignment of subjects into alternative treatments, and there being no spillover effects across treatments, we can then test for *individual* violations of the axioms of revealed preference through pairwise comparisons of the behavior of a given recipient

[1] Following Harrison and List [2004], the experiment falls into the 'natural field' category as the subjects naturally undertake the task in the environment under study and they are not aware of being involved in an experiment.

in the treatment they were actually assigned to, to their predicted behavior in counterfactual treatments.

Our main result is that on both the extensive and intensive margins of charitable giving, individual choices can be rationalized within a standard model of consumers maximizing utility subject to budget constraints, where individual preferences are defined over own consumption and charitable donations received by the project. The behavior of at least 80% of recipients who make some positive contribution is in line with their preferences satisfying GARP. We cannot reject that a continuous, convex, and monotone utility function could have generated their choices. In short, in a real world environment where participants make simple decisions they are familiar with, the predictions of microeconomic theory work well in explaining individual behavior.

We view our methodological approach using a natural field experiment as complementary to non-experimental tests of consumer theory which typically exploit panel data on consumer purchases. However, as with within-subject experimental designs, in non-experimental data apparent violations of revealed preference might instead be due to changes in tastes, changes in the holding of durables, or the storage of consumables. Moreover, non-experimental methods suffer from consumption expenditures being measured with error, and there being stringent data requirements for information on labor supply or asset demands, as well as consumption expenditures.

Consumer panels also typically suffer from observed price changes being both relatively small, and not necessarily implying an intersection of budget sets. Hence in contrast to our research design, tests of revealed preference based on non-experimental data are likely to have low power [Varian 1982, Bronars 1995]. Such approaches, have provided mixed results with some studies rejecting behavior consistent with GARP [Koo 1963, Mossin 1972, Hardle *et al* 1991] and others finding more rationalizable patterns of consumption [Manser and McDonald 1988, Famulari 1995]. Recent methodological advances using non-parametric techniques suggest consumer behavior does not reject GARP in the long run for most income groups [Blundell *et al* 2003].

Our analysis also builds on laboratory evidence on consumer choice, which has provided mixed evidence on whether individual behavior is consistent with GARP [Battaglio 1973, Cox 1997, Sippel 1997, Harbaugh *et al* 2001, Andreoni and Miller 2002, List and Millimet 2005, Choi *et al* 2007]. Our research design combines the key advantages of laboratory experiments in being able to experimentally manipulate the economic environment faced by agents and requiring relatively weak data requirements, with the advantages of a field study using real world data on a large population. As suggested by Varian [2006], this research design is perhaps the best possible that could be used to test whether individual behavior is consistent with revealed preference theory.[2]

[2] Our results differ from some of the laboratory evidence on consumer choice, such as Battalio *et al* [1973] and Sippel [1997] who find behavior not to be in line with GARP. This may be because in our study consumers are faced with a real life setting and make simple decisions they are familiar with, and we exploit a large sample of individuals. In Sippel's experimental design, subjects were first confronted with various budget sets defined over a number of goods such as magazines, video clips, soft drinks, and snacks. For each budget set subjects had to choose a bundle of goods they wanted to consume afterwards. One of the budget sets was then randomly drawn and subjects received the goods they had chosen for post-experimental consumption. Battalio *et al*'s study is based on a token economy using data from 37 individuals in a mental hospital. An additional reason why a field

Finally, we reiterate that our analysis here focuses on the broad question of whether individual behavior is consistent with neoclassical microeconomic theory. In a companion paper we exploit the natural field experiment to shed light on specific issues relating to the economics of charitable giving. In particular, in Huck and Rasul [2007] we build on evidence from other recent natural field experiments on charitable giving [Eckel and Grossman 2006, Karlan and List 2007] to provide evidence on the relative efficacy of the alternative fundraising schemes in our treatments.

The paper is organized as follows. Section 2 describes the natural field experiment, and presents a conceptual framework of consumer choice from which to understand behavior across the treatments. Section 3 provides descriptive evidence on the extensive and intensive margins of charitable giving in each treatment, and tests whether, *on average*, the behavior of recipients in each treatment is in accordance with revealed preference theory. Section 4 presents the econometric method and results that tests directly for *individual* violations of revealed preference theory. Section 5 concludes.

2 The Natural Field Experiment

2.1 Design

In June 2006 the Bavarian State Opera organized a mail out of letters to over 25,000 individuals designed to elicit donations for a social youth project the opera was engaged in, "Stück für Stück". The project's beneficiaries are children from disadvantaged families whose parents are almost surely not among the recipients of the mail out. Hence the fundraising campaign relates to a project that conveys no immediate benefits to potential donors. As such, the campaign is more similar to fundraising by aid charities, rather than the typical forms of opera fundraising where money is collected for projects that benefit opera attendees directly. The experimental design allows us to focus on whether consumer behavior is consistent with a standard model of consumer choice because it rules out any role for two types of non-standard behaviors that have been relevant in other studies of charitable giving. In particular, there is no role for social recognition to drive donor behavior, as in Andreoni and Petrie [2004], as donors are not publicly announced.

The project finances small workshops and events for schoolchildren with disabilities or from disadvantaged areas. These serve as a playful introduction to the world of music and opera. It is part of the Bavarian State Opera's mission to preserve the operatic art form for future generations and the project is therefore a key activity to fulfill this mission. As it is not one large event that donations are sought for, but rather a series of several smaller events, it is clear to potential donors that additional money raised can fund additional activity. In other words, the marginal contribution will *always* make a difference to the project.

experiment may be preferable to laboratory evidence when examining the choice between a charitable good and all other private consumption, is that field and laboratory evidence on behavior with regards to charitable giving can conflict in these different environments [Rondeau and List 2008].

The original 25,000 mail out recipients were randomly selected from the Opera's database of 50,000 customers who had bought at least one ticket to attend either the opera or ballet, in the twelve months prior to the mail out. The database also contains information on each individual's attendance at the opera, the number of tickets purchased, the number of separate ticket orders, the value of tickets purchased, and zip code of residence. We remove non-German residents, corporate donors, formally titled donors, and recipients to whom we cannot assign a gender—typically couples. We also remove recipients that were assigned to one treatment that is not relevant to this study but that is exploited in Huck and Rasul [2007]. The working sample is then based on the remaining 18,725 individuals and the analysis throughout refers to this sample.

The mail out recipients were randomly assigned to one of five treatments that varied in how individual donations would be matched by an anonymous lead donor. The precise format and wording of the mail out is provided in the Appendix. The mail out letters were identical in all treatments with the exception of one paragraph. As explained in detail in the next subsection, individuals assigned to each treatment effectively face different budget sets defined over their contribution towards the charitable good and their own consumption.[3]

Donations are matched by an anonymous lead donor. Since the presence of the lead donor may serve as a signal of the underlying quality of the project Andreoni [2006a], it is essential that the lead donor is also mentioned in a non-matching baseline treatment. Hence the control treatment was such that recipients were only informed that the project had already garnered a lead gift of €60,000, and there is no offer to match donations—a donation of one Euro corresponds to one Euro being received for the project. The wording of the key paragraph read as follows,

T1 (Control): *A generous donor who prefers not to be named has already been enlisted. He will support "Stück für Stück" with €60,000. Unfortunately, this is not enough to fund the project completely which is why I would be glad if you were to support the project with your donation.*

This paragraph is manipulated in the other treatments. The next two treatments provided recipients with the same information on the presence of a lead donor as in the control treatment, but varied the rate at which donations would be matched. Individuals assigned to these treatments effectively face a lower relative price for the charitable good vis-à-vis own consumption, than do individuals in the control treatment. The first matching treatment informed recipients that each donation would be matched at a rate of 50%, so that giving one Euro would correspond to the opera receiving €1.50 for the project. The corresponding paragraph in the letter read as follows,

T2 (50% Matching): *A generous donor who prefers not to be named has already been enlisted. He will support "Stück für Stück" with up to €60,000 by donating, for each Euro that we receive within the next four weeks, another 50 Euro cent. In light of this unique opportunity I would be*

[3] All letters were designed and formatted by the Bavarian State Opera's staff, and addressed to the individual as recorded in the database of attendees. Each recipient was sent a cover letter describing the project, in which one paragraph was randomly varied in each treatment. On the second sheet of the mail out further details on the "Stück für Stück" project were provided. Letters were signed by the General Director of the opera house, Sir Peter Jonas, and were mailed on the same day—Monday 19th June 2006.

glad if you were to support the project with your donation.

The next treatment was identical to T3 except the match rate was set at 100%, so the corresponding paragraph in the mail out letter was,

T3 (100% Matching): *A generous donor who prefers not to be named has already been enlisted. He will support "Stück für Stück" with up to €60,000 by donating, for each donation that we receive within the next four weeks, the same amount himself. In light of this unique opportunity I would be glad if you were to support the project with your donation.*

The fourth treatment presented recipients with a match rate of 100% conditional on the donation given being above a fixed threshold—€50. Below the €50 threshold the match rate was zero. Hence individuals effectively faced a non-convex budget set defined over the donation received by the charitable project and their own consumption. This matching scheme was explained in the mail out letter as follows,

T4 (Non-convex): *A generous donor who prefers not to be named has already been enlisted. He will support "Stück für Stück" with up to €60,000 by donating, for each donation above €50 that we receive within the next four weeks, the same amount himself. In light of this unique opportunity I would be glad if you were to support the project with your donation.*

Note that in this treatment, for donations below €50 recipients have the same information on the lead donor and face the same relative price of giving, as in the control treatment T1. For donations above €50 recipients have the same information on the lead donor and face the same relative price of giving as in the 100% matching treatment, T3. This generates a series of revealed preference predictions. The non-convexity in the budget set might potentially also introduce a focal or reference point for donations at € 50. If such reference points influence behavior, then recipients who would have otherwise given at least €50 under treatment T3, might be pushed to reduce their donation given towards €50 under T4 and thus violate the predictions of revealed preference. We later test for such violations.

The final treatment offered recipients a fixed positive match of €20 for *any* positive donation. This corresponds to a parallel shift out of the budget line that is similar to a pure income effect leaving the relative price of donations received unchanged. We refer to this treatment as the 'leverage' treatment because any donation slightly above zero has an extremely high implicit matching rate. This treatment was explained in the mail out letter as follows,

T5 (Leverage): *A generous donor who prefers not to be named has already been enlisted. He will support "Stück für Stück" with up to €60,000 by donating, for each donation that we receive within the next four weeks regardless of the donation amount, another €20. In light of this unique opportunity I would be glad if you were to support the project with your donation.*

The induced budget constraint in this treatment intersects those of the other matching treatments which generates a further set of revealed preference predictions.

There are two points to bear in mind regarding the design of the experiment. First, the opera house had no explicit target in mind to be raised, nor was any such target discussed in the mail out. Second, recipients are told the truth—the lead gift was actually provided and each matching scheme implemented. The value of matches was capped at €60,000 which ensured subjects were told the truth even if the campaign was more successful than anticipated. If recipients have the same belief that others had donated to such an extent that the €60,000 of the lead donor was already exhausted and so the match would no longer be in place, there should be no difference in behaviors across treatments. This hypothesis is firmly rejected by the data.[4]

2.2 Conceptual Framework

We assume potential donors have preferences defined over two dimensions—their own consumption and the marginal benefit their donation provides. Given the linearly expandable nature of the project, this marginal benefit relates to donation received by the project, d_r. In our setting we then have two goods—donations received by the project, and a composite good representing all other consumption. We denote the price and goods vectors as $\mathbf{p}$ and $\mathbf{x}$ respectively. As in the exposition of Varian [2006], we then have the following definitions.

Definition (Revealed Preference) Given some vector of prices and chosen bundles $(\mathbf{p}^t, \mathbf{x}^t)$ for $t = 1, ..., T$, $\mathbf{x}^t$ is directly revealed preferred to $\mathbf{x}$ if $\mathbf{p}^t\mathbf{x}^t \geq \mathbf{p}^t\mathbf{x}$. $\mathbf{x}^t$ is revealed preferred to $\mathbf{x}$ if there is some sequence $r, s, t, ..., u, v$ such that $\mathbf{p}^r\mathbf{x}^r \geq \mathbf{p}^r\mathbf{x}^s$, $\mathbf{p}^s\mathbf{x}^s \geq \mathbf{p}^s\mathbf{x}^t, ..., \mathbf{p}^u\mathbf{x}^u \geq \mathbf{p}^u\mathbf{x}$.

Definition (Weak Axiom of Revealed Preference) If $\mathbf{x}^t$ is directly revealed preferred to $\mathbf{x}^s$, then it is not the case that $\mathbf{x}^s$ is directly revealed preferred to $\mathbf{x}^t$, so that $\mathbf{p}^t\mathbf{x}^t \geq \mathbf{p}^t\mathbf{x}^s$ implies $\mathbf{p}^s\mathbf{x}^s < \mathbf{p}^s\mathbf{x}^t$.

Definition (Generalized Axiom of Revealed Preference) The data $(\mathbf{p}^t, \mathbf{x}^t)$ satisfy the Generalized Axiom of Revealed Preference (GARP) if $\mathbf{x}^t$ is revealed preferred to $\mathbf{x}^s$ implies $\mathbf{p}^s\mathbf{x}^s \leq \mathbf{p}^s\mathbf{x}^t$.

In two dimensions as in our setting, the Weak and Generalized Axioms of Revealed Preference are equivalent [Rose 1958]. The main result in the revealed preference literature is from Afriat [1967] which states,

Afriat's Theorem Given some choice data $(\mathbf{p}^t, \mathbf{x}^t)$ for $t = 1, ..., T$, the following conditions are equivalent:

(i) The data satisfy GARP.

[4] In treatments T2 to T5, recipients were told the matching schemes would be in place for four weeks after receipt of the mail out. This deadline was not binding—over 97% of recipients that donated did so during this time frame and the median donor gave within a week of the mail out. Moreover we find no evidence of differential effects on the time for donations to be received between any treatment and the control treatment, in which no such deadline was announced. There might also be a slight concern that if recipients know each other, then having knowledge of whether another opera attendee had received the mail out, and the form of the letter they received, this might in principle lead to some changes in behavior if there are strong peer effects in charitable giving. We note however, that the opera house received no telephone queries regarding treatment differences.

(ii) There exists a non-satiated, continuous, monotone, and concave utility function, $u(\mathbf{x})$ that rationalizes the data.

In our setting, this corresponds to allowing each individual's behavior to be rationalized as the following utility maximization problem,

$$\max_{d_r} u(c, d_r) \text{ subject to } c + d_g \leq y,\ c, d_g \geq 0, \text{ and } d_r = f(d_g), \tag{1}$$

where $u(c, d_r)$ has the properties listed above, the first constraint ensures consumption can be no greater than income net of any donation given, $y - d_g$, the second constraint requires consumption and donations given to be non-negative, and the third constraint denotes the matching scheme that translates donations given into those received by the opera house. Under linear matching treatments for example, $d_r = \lambda d_g$.

Figure 1 graphs the budget sets induced by the five treatments in $(y - d_g, d_r)$-space. Each treatment induces the consumer to face different price vectors. In the control treatment (T1) the budget line has vertical intercept y and a slope of minus one as for each Euro given by an individual, the project receives one Euro ($d_r = d_g$). Treatments T2 and T3 vary the price of donations relative to own consumption so that with the 50% match rate in T3, $\lambda = 1.5$, and with the 100% match rate in T4, $\lambda = 2$. In both cases the budget set pivots out with the same vertical intercept. Treatment T4 introduces a non-convex budget constraint that partly overlaps with those in T1 and T3. In this treatment, as the match rate λ is a function of d_g, there are kinks in the budget line that might lead to an interior corner solution in the individual optimization problem above. Treatment T5 causes a parallel shift out of the budget line by €20 of for any positive donation, as if consumer income had exogenously increased.

As the budget sets in treatments T1 to T5 intersect, pairwise comparisons of the behavior of individuals in any two treatments allows us to test whether consumer behavior is on average consistent with GARP. However, although behavior on average might be consistent, each individual's preferences may violate GARP. We therefore exploit the random assignment of recipients to treatments to test for *individual* violations of GARP. The research design provides a series of such tests based on pairwise comparisons of the behavior of a given recipient in the treatment they were actually assigned to, to their predicted behavior in counterfactual treatments.

3 Descriptive Evidence

3.1 Sample Characteristics and Treatment Assignment

Individuals were randomly assigned to one of the five treatments. Table A1 summarizes the individual information from the database. For each observable, Table A1 reports the p-values on the null hypothesis that the mean characteristic of individuals in the treatment group are the same as in the control group T1. There are no significant differences along any dimension between recipients in each treatments, so that individuals are indeed randomly assigned into treatments.

Columns 1 and 2 show that there is an almost equal split of recipients across treatments, and that close to half the recipients are female. Columns 3 to 7 provide information on individuals' attendance at the opera as measured by the number of tickets the individual has ordered in the twelve months prior to the mail out, the number of separate ticket orders that were placed over the same period, the average price paid per ticket, and the total amount spent. Individuals in the sample typically purchase around six tickets in the year prior to the mail out in two separate orders. The average price per ticket is €86 with the annual total spent on attendance averaging over €400. We use information on the zip code of residence of individuals to identify that around 40% of recipients reside within Munich, where the opera house is located. Column 8 provides information on the proportion of individuals that bought a ticket in 2006, namely in the six months prior to the mail out. The majority of individuals have attended the opera in 2006.

Recipients are not representative of the population—they attend the opera more frequently than the average citizen and are likely to have higher disposable incomes. Our analysis therefore sheds light on how such selected individuals donate towards a project that is being directly promoted by the opera house. To the extent that other organizations target charitable projects towards those with high affinity to the organization as well as those who are likely to have high income, the results have external validity in other settings. Moreover, while the non-representativeness of the sample may imply the observed *levels* of response or donations likely overstate the response among the general population, we focus attention on *differences* in behavior across treatments that purge the analysis of the common characteristics of sample individuals. We return to discuss in more detail the external validity of our results in the final section.

3.2 Recipient Behavior: Extensive and Intensive Margins

Table 1 provides descriptive evidence on behavior on the intensive and extensive margins of charitable giving by treatment. For each statistic we report its mean, its standard error in parentheses, and whether it is significantly different from that in the control treatment. Figure 1 provides a graphical representation of the outcomes across treatments, showing for each treatment t the average bundle chosen, $\mathbf{x}^t$, at the relevant price vector, $\mathbf{p}^t$. Among the full sample of 25,000 recipients more than €120,000 were donated, fully exhausting the €60,000 of the lead donor. In our working sample of 18,725 individual recipients, Columns 1 to 3 reveal that in aggregate 780 individuals donated a total of €75,350, which corresponds to €116,489 actually raised for the project, with a mean donation given of €96.6.[5]

Column 4 shows that response rates vary from 3.5% to 4.7% across treatments, which are

[5]This exceeded the expectations of the Bavarian State Opera which were that there would be a 1% response rate and mean donation of €100. The total amount raised further exceeded the opera house's expectations because the figures we report for our analysis excludes donations from non-German residents, corporate donors, formally titled donors, and recipients that correspond to a husband and wife so that no gender can be assigned.

almost double those in comparable large-scale natural field experiments on charitable giving [Eckel and Grossman 2006, Karlan and List 2007]. However, despite there being large variations in the budget sets consumers face in treatments T1 to T3, there are no significant differences in response rates across these treatments. Although the point estimates of response rates are higher in the linear matching treatments, the data suggests changes in the relative price of charitable giving do not significantly affect behavior along the extensive margin. Hence there are few consumers who are just on the margin of donating in treatments T2 or T3, namely those for whom the marginal rate of substitution between own consumption and donations received, is such that $-\frac{1}{2} < MRS_{c,d_r}|_{d_r=0} < -1$.[6]

However, as made clear in the conceptual framework discussed earlier, treatment T5—that causes a parallel shift out of the budget set for any positive donation—is the treatment that should induce the largest change in the number of donors relative to the control group. This is because standard consumer theory predicts any individual with preferences such that $MRS_{c,d_r}|_{d_r=0} < 0$ will find it optimal to donate some amount in T5, whereas this is not the case in other treatments. The data supports this—the response rate is significantly higher in T5 relative to the other treatments. However, the fact that the response rate in T5 is 4.7% highlights that even among this targeted population, 95% of individuals cannot be induced to donate. These individuals either do not value the project at all, or, face transactions costs that are sufficiently high to offset any warm glow they feel from giving to this particular cause.

On the intensive margin, Column 5 shows that in the control treatment T1, the average donation given is €132. As the relative price of donations received falls in treatments T2 and T3, the average donation received increases to €151 in T2 with a 50% match rate, and to €185 in T3 with a 100% match rate. Importantly, as shown in Figure 1 and Column 7 of Table 1, as the match rate increases, the average donation *given*, d_g, falls from €132 in the control treatment T1 to to €101 in T2 with a 50% match rate, and to €92.3 in T3 with a 100% match rate. Column 8 reiterates that these differences are not driven by outliers—the median donation given is significantly lower in treatments T2 and T3 than the control treatment T1.[7]

Treatment T4 induces recipients to face a non-convex budget set. For donations below €50 the budget line is coincident with that of the control treatment T1, for donations at or above €50 it coincides with that of the 100% matching treatment T3. Figure 1 shows that average outcome in terms of donations given and received in T4 replicate almost exactly those in the 100% matching

[6] One explanation for the high response rates we obtain may be that the Bavarian State Opera has not previously engaged in fundraising activities through mail outs, nor is the practice as common in Germany as it is in the US. Another explanation—which would have more significant implications for external validity—is that the mail out took place during late June and early July 2006 during a period when Germany was hosting the soccer world cup. Dohmen *et al* [2006] provide evidence that individual perceptions and expectations of Germans are altered significantly one day after each of the seven matches played by the German national team. However we find no empirical relationship between such dates and the likelihood individuals donate nor the amount donated.

[7] An immediate implication is that such linear matching schemes lead to a partial crowding out of donations given, and so do not pay for the fundraiser overall. This is explored in greater depth in Huck and Rasul [2007] but is not of direct relevance to the analysis here on testing revealed preference theory.

treatment T3—the average donation received in T4 is €194, as opposed to €185 in T3, and the average donation given is €97.9, as opposed to €92.3 in T3. To see why this is so, note that in the control treatment the average donation received is €132. This suggests the portion of the budget line in T4 that lies to the left of €100 on the x-axis of donations received is irrelevant for many recipients. In essence, treatments T3 and T4 present the average recipient with an almost identical choice. Hence response rates and donations should not differ markedly between the two.

Finally, comparing the leveraged treatment T5 to the control treatment, as previously discussed, response rates are significantly higher in T5 than in T1. Consumer theory also suggests these additional donors should be willing to contribute relatively low amounts to the project. This is strongly supported in the data, as best illustrated in Figure 1—there is a decrease in both the donations given and received in treatment T5 relative to the control group. Column 5 of Table 1 shows the average donation received in T5 is €89.2—relative to T1, donations given fall by significantly more than €20. This result is not driven by outliers. Column 6 shows the median donations received is also significantly lower (by €30) in T5 than T1. These effects remain even in Columns 7 and 8 when differences in the mean and median amounts *given* are considered.

At face value this suggests pure income transfers lead to more than full crowding out of donations given so that charitable giving is an inferior good. While this does not of course violate the axioms of revealed preference, this interpretation is however misleading. To understand more precisely how small changes in individual wealth affect charitable giving, we need to take account of the change in composition of donors in treatment T6 relative to treatment T2. Hence in the later econometric analysis we estimate changes in behavior conditional on the observable characteristics of donors.[8]

3.3 Violations of Revealed Preference in Aggregate

As the budget sets in treatments T1 to T5 intersect or overlap as shown in Figure 1, pairwise comparisons of the average behavior of individuals in any two treatments lead to tests of whether behavior is consistent with revealed preference theory. These tests are of three types and relate to—(i) the proportion of recipients that should donate some positive amount; (ii) the proportion of recipients that lie above or below some critical threshold, which is typically where the two budget lines intersect; and, (iii) the distribution of donations given and received.

An example of the first type of test is given by comparing treatments T1 and T3. As shown in Figure 1, the budget set expands moving from T1 to T3. Assuming individual preferences are well behaved, the proportion of individuals that find it optimal to provide some positive donation under T3 should be at least as great as the proportion that respond under T1.

An example of the second type of test is given by comparing treatments T2 and T5 in which

[8] To reiterate, the data does support the hypothesis that recipients had homogenous beliefs that others had donated to such an extent that the €60,000 of the lead donor was already exhausted and so the match scheme would no longer be in place. If so, there should be no difference in behaviors across treatments.

the budget sets cross at donations given equal to €40. For all donations given greater than €40, the budget set expands under T2 relative to T5. Hence revealed preference arguments imply the proportion of donations given that are at least €40 should be weakly higher in T2 than T5.

An example of the third type of test is given by comparing treatments T3 and T4. As shown in Figure 1, the budget sets are coincident for donations given that are more than €50. Hence the distribution of donations given conditional on them being more than €50, should be identical in both treatments. This follows from the fact that any donors that contribute strictly more than €50 under T3 should, by revealed preference, also contribute the same under T4.

Table 2 presents the full set of such tests for each pairwise comparison of treatments. One test—relating to the distribution of donations given—is boxed as it requires the additional assumption of strict convexity in addition to satisfying GARP. For each test, we report the p-value on the null hypothesis consistent with revealed preference. Of the twelve tests, the results of eleven do not contradict consumers, on average, having an underlying utility function that displays standard properties.

The one exception relates to the test between treatments T3 and T4 reported in the last column. To examine this violation in more detail we note that if preferences are convex, then by revealed preference, individuals who would have donated less than €50 in T3 either ought to choose optimally not to donate at all in T4, give strictly less than €50, or choose the interior corner solution and give exactly €50. Hence relative to T3, there ought to be relatively more donations given *below or at* $d_g =$ €50 in T5.

To check for this, Figure 2A shows the difference in the proportion of donations given in €10 bins from €0 and €199, between treatments T4 and T3. The most striking fact that emerges is bunching in the distribution of donations given in T4 relative to T3 slightly above $d_g =$ €50, and the fall in the proportion of donations given below €50. We find that 20.6% of respondents choose to donate between €51 and €60 in treatment T4, whereas there are zero contributions in this range in T3. These individuals are *not* those who otherwise would have given more than €60—the two distributions are almost identical for donations above €70. Rather, individuals that would have given *less* than €50 in T3 are those now giving close to €60 under T4. Hence we find that donors prefer to give incrementally above €50 when faced with the non-convex budget set—while 28.3% of respondents in T3 choose to give €50, this actually falls to 13.1% in T4.

This may reflect there being small psychological costs of donating the exact amount mentioned in the mail out letter—perhaps, donors do not want to 'exploit' the generous lead donor by giving the minimum amount required for the match to be implemented. In any case, there is no evidence of consumers being swayed by a focal point effect as there is no bunching from above at $d_g =$ €50 in T4 relative to T3. In short, the reported violation of revealed preference theory reported in the last column of Table 2 is entirely due to recipients not moving precisely to the interior corner solution at €50 in T4 but rather overshooting by €10 or so. If we adjust for this effect and test whether the distributions in T3 and T4 above $d_g =$€70 are identical, we cannot reject the null.

Taken together, the evidence suggests revealed preference theory performs remarkably well in predicting aggregate consumer behavior in our study. Although on average, consumer behavior does not violate GARP, each and every individual alone may still violate GARP. Therefore in Section 4 we exploit the twin identifying assumptions of random assignment of subjects into treatments, and there being no spillover effects across treatments, to provide evidence on *individual* violations of GARP by comparing the actual choices of consumers in treatments they are assigned to, to their predicted choices in any counterfactual treatment.

4 Econometric Evidence

4.1 Empirical Method

The research design allows us to test whether behavior is in line with GARP along both the extensive and intensive margins of charitable giving. To begin with, we define a dummy variable, D_i, equal to one if individual i donates to the charitable project, and equal to zero otherwise. To shed light on the extensive margin of charitable giving, we estimate the following equation using a probit model for whether any donation is given or not,

$$\textbf{prob}\,(D_i = 1) = \textbf{prob}\,(u_i > -(\beta_1 T_i + \gamma_1 X_i))\,. \tag{2}$$

Whether i donates or not depends on the budget set she faces as embodied in the treatment she is assigned to, T_i. Given random assignment this is orthogonal to the error term u_i so that $\widehat{\beta}_1$ provides a consistent estimate of the treatment effect on the extensive margin of giving of being assigned to treatment T_i relative to whichever is the omitted treatment. We control for individual characteristics X_i, to reduce the sampling errors of the treatment effect estimates. We report marginal effects and calculate robust standard errors throughout.

On the intensive margin of charitable giving, the central econometric concern is that even with random assignment into treatments, we cannot in general make valid causal inferences conditional on donations being positive because those that choose to donate are likely to differ from those that choose not to donate. We address this sample selection issue in two ways. First, we estimate for the entire sample of recipients the following OLS specification for the donation received by the opera house from recipient i, d_{ri},

$$d_{ri} = \beta_2 T_i + \gamma_2 X_i + v_i, \tag{3}$$

so that $d_{ri} = 0$ for non donors, v_i is a disturbance term and all other controls are as previously defined. We calculate robust standard errors throughout. Under the assumption of no spillover effects between treatments, the parameter of interest β_2 then measures the average treatment effect on the donation received of individual i being assigned to treatment T_i relative to whichever is the omitted treatment.

Inference can be made on positive donations only under the condition that $E[v_i|T_i, X_i, D_i = 1] = 0$ [Angrist 1997]. Intuitively, we require unobserved determinants of the amount donated, conditional on giving, treatment assignment and observable characteristics, to be orthogonal to unobserved determinants of the decision to donate. We provide suggestive evidence in support of this assumption using results from a follow-up field experiment with the same opera house [Huck and Rasul 2008].

In particular, a month after the original mail outs, recipients that did not respond were sent a reminder letter. This stated that the original matching scheme, if any, was no longer in place, but still encouraged individuals to donate to the project. We then consider the response to the reminder letter of those individuals that did not respond to the original mail out letter, to shed light on the nature of selection into response.

First, we find there are individuals that donate following the reminder even though they did not respond to the original mail out. The response rate to the reminder was 1.8%. If only individuals with the highest valuation of the charitable good ever respond to mail outs, then the response to the reminder ought to be zero. This suggests individuals face transaction costs to responding.

Second, the average donation under the reminders was €78.7. This is not significantly different from the €74.3 that was donated in the original control treatment T1. This suggests the transactions costs faced by recipients are orthogonal to their valuation of the charitable good, so that unobserved determinants of how much to donate may be less correlated to the underlying transactions costs faced that determine whether to give or not. While not conclusive, the data is suggestive of decisions to donate being driven by the existence of transactions costs that are unrelated to how much individuals would like to donate.

In order to estimate the treatment effects conditional on a positive donation being made, we specify a hurdle model which takes explicit account of the fact that the initial decision to donate ($D_i = 0$ or 1) may be separated from the decision of how much to donate, namely, the choice over d_r conditional on $D_i = 1$. A simple two-tiered model for charitable giving has, as a first stage, the probit model above in (2). At the second stage, we assume donations received are log normally distributed conditional on any donation being given, namely, $\log(d_{ri})|(T_i, X_i, D_i = 1) \sim N(\beta_3 T_i + \gamma_2 X_i, \sigma^2)$. The maximum likelihood estimator of the second stage parameters, (β_3, γ_2), is then simply the OLS estimator from the following regression,

$$\log(d_{ri}) = \beta_3 T_i + \gamma_3 X_i + z_i \text{ for } d_{ri} > 0, \tag{4}$$

where we calculate robust standard errors throughout [Wooldridge 2002]. For each treatment, we therefore present both the OLS and hurdle model estimates, with the caveats associated with each set of results.[9]

[9] In theory, we could use non-parametric techniques to also deal with these selection issues. Chay and Powell [2001] provide an overview of these methods.

The individual characteristics controlled for in X_i are whether recipient i is female, the number of ticket orders placed in the 12 months prior to mail out, the average price of these tickets, whether i resides in Munich, and a dummy for whether the year of the last ticket purchase was 2006.[10]

Finally, we note that the error terms in each specification above capture factors that influence the propensity to donate, or the level of donations. Such factors may include reciprocity, guilt, or shame for example [Andreoni 2006b]. One underlying identifying assumption is that such factors do not differentially affect behavior across the treatments and so valid inference can still be made on the change in behavior moving from one treatment to another.

4.2 Results: Recipient Behavior

4.2.1 Extensive Margin

Table 3 presents the estimates of the parameter of interest β_1 from the probit model (2) on the intensive margin of charitable giving. To begin with we note that on the extensive margin, in line with the descriptive evidence, changes in the relative price of charitable giving do not significantly affect the likelihood that recipients donate (Column 1).

In terms of the non-convex treatment, recipients are slightly more likely to respond to this treatment than to the control treatment (Column 2a). Using treatment T3 as the comparison group as the two budget lines coincide for donations given greater than or equal to €50, there is again no evidence of response rates being higher in T4 than T3 (Column 2b).

In terms of the leveraged treatment, recipients are significantly more likely to respond to this treatment than to the control treatment (Column 3a), in line with revealed preference theory. Quantitatively, recipients are 1.3% more likely to respond to this treatment than the control treatment. The magnitude of this effect is large given that the response rate in the control group of 3.5%. To purge this estimate of donors that are likely to have only contributed to T5 and not to T1, the next specification restricts the sample to those that, *a priori*, are expected to donate larger amounts. To define such donors we use data only from the control treatment and run an OLS regression of donations given on the individual characteristics in X_i described above. We therefore form a predicted donation for each and every recipient in the data base and define 'high valuation' donors to be those predicted to give more than the mean donation based on these observables. Reassuringly, the result in Column 3b shows there are no significant differences in response rate in this subsample, suggesting that new donors that only enter under T5 are likely to donate small amounts as predicted by standard theory.

[10] We also experimented with alternative controls in X_i. For example, the number of tickets bought may serve as an alternative proxy for affinity rather than the number of orders placed. However, we prefer the latter measure as the former may be confounded by recipients attending the opera with their friends and family. In any case, the main results are robust to slight alterations in the controls.

4.2.2 Intensive Margin

Table 4 presents the estimates of the parameters of interest (β_2, β_3) from the OLS (3) and hurdle (4) models on the extensive margin of charitable giving. The first column focuses on the effect on individual's choice bundle of own consumption and donations towards the charitable good, as the relative price of the charitable good changes, as induced by linear match rates in treatments T2 and T3. To begin with we see that as the match rate increases from zero to 50% (T1 to T2), the OLS estimates show that larger donations are received in treatment T2, although this is not quite significantly different from zero at conventional levels. It is however the case that as the match rate increases to 100% in treatment T3, significantly larger donations are received relative to the control group (Column 1a). These results are qualitatively replicated using the second stage hurdle model estimates from (4) in Column 1b. The similarity between the OLS and hurdle model estimates suggests unobserved determinants of the amount donated, conditional on giving, treatment assignment and observable characteristics, are orthogonal to unobserved determinants of the decision to donate.

In terms of the non-convex treatment T4, the OLS estimates imply significantly larger donations are received in this treatment relative to the control treatment T1 (Column 2a), a result confirmed using the hurdle model estimates (Column 2b). Similarly, comparing the choice bundles when individuals face the non-convex budget set to when they are offered a 100% match rate in T3—when a portion of these budget lines coincide—the donation received is higher in T4 than in T5 (Columns 2c and 2d).

Given the previous results show response rates to be no lower in T4 vis-à-vis T1, this is in line with standard theory that predicts those that would have donated less than €50 in T1 should find it optimal to move towards the interior corner solution and donate €50 under T4. As treatments T3 and T4 present the average recipient with an almost identical choice if they donate more than €50, Columns 2e and 2f repeat the exercise for the subset of recipients who are, *a priori*, expected to be among the highest donors. More precisely, to define such donors we use data only from the control treatment to run an OLS regression of donations given on the individual characteristics in X_i described above, and to therefore form a predicted donation for each and every recipient in the data base. We then define such 'very high valuation' donors to be those predicted to give in the top quintile of donations based on these observables. In this subsample donations given are not significantly different in T4 and T3—using either the OLS or hurdle model estimates. This is as expected given that these recipients face essentially the same choice set.

Finally, in terms of the leveraged treatment, we see that, in contrast to the descriptive evidence, once observables are conditioned on—(i) there are no significant differences in the donations received between the leverage and control treatments using either the OLS or hurdle model estimates (Columns 3a and 3b); (ii) this remains the case when we restrict attention to high valuation donors (Columns 3c and 3d) who would be expected to have donated more than €20 in the absence of the offered match (Columns 3c and 3d). Hence when faced a matching scheme that essentially provides a small income transfer to individuals, they reduce one-for-one their donations given to leave unchanged the donations actually received by the charitable project. The results suggest the leveraged treatment increases the proportion of recipients that donate, but for those that do, the offered match leads individuals to reduce their donation given to such an extent that the donations actually received by the project are left unchanged overall. Hence, as explored more fully in Huck and Rasul [2007], there is a full crowding out of donations given when individuals are provided small income transfers as in treatment T5.

4.3 Results: Individual Violations of Revealed Preference

In our between-subject research design we do not observe the same consumer making multiple choices under alternative budget sets. To therefore detect *individual* violations of GARP we estimate for each individual, whose actual choice we only observe in treatment t, what she would have donated in the relevant counterfactual treatment $t' \neq t$ based on the predictions from the hurdle model. This gives two decisions per individual—one actually observed, one predicted—which we use as the basis of tests for individual violations of revealed preference theory. The counterfactual treatment t' is chosen so that GARP makes a prediction on how behavior in treatment t and t' are related. In our research design there are 10 such pairwise comparisons, as shown in Table 5.

The method is valid under the twin identifying assumptions of random assignment of subjects into alternative treatments, and there being no spillover effects across treatments [Heckman *et al* 1999]. Table A1 already suggested subjects were randomly assigned into treatments, and there is little evidence of there being spillover effects across treatments.[11] As a first step we therefore use our estimates from (4) to map observed characteristics into choice bundles, $\mathbf{x}^t$. For an individual whom we observe in treatment t we then obtain predicted donations for the relevant comparison treatment t' by substituting their characteristics into the estimated models.

Each comparison is then made among the subset of recipients that donated some positive amount in whichever treatment they were assigned. This avoids us having to impute an expected donation to those 95% of recipients that are unlikely to donate under any circumstances. Moreover, as Table A2 shows, the observable characteristics of those that donate in their assigned treatment, do not differ significantly from those that donate in the control treatment. This is true across all treatments in every observable dimension, both in terms of their mean values, and the overall distribution of these observables. Using our approach we are thus able to replicate the within-subject experimental research designs of Sippel [1997] and Andreoni and Miller [2002], in which each subject makes choices under multiple treatments.

Table 5 then presents the various tests for individual violations of revealed preference that can be derived from pairwise treatment comparisons. These are analogous to a subset of the tests for the average consumer behaving in accordance with GARP documented in Table 2. Column 1 shows the number of violations of revealed preference theory for each pairwise comparison of

[11] For example, there were no cases of individuals calling the opera house enquiring about alternative treatments.

treatments. We also show the proportion of violations defined as the number of violations divided by the total number of positive donations given in the treatment from which actual (and not predicted) donations are used. Both measures have been previously used in the literature as measures of goodness of fit in tests of revealed preference. For example, Whitney and Swofford [1987] use the former measure, and Famulari [1987] uses the latter.[12]

Across the pairwise comparisons, the proportion of violations of revealed preference theory varies from zero to 20%. For example, there are three violations arising from a comparison of observed behavior in T1 and predicted behavior in T4, so that 2.3% of all donors that actually made a positive donation under the control treatment T1 are predicted to make a donation that violates GARP when faced with the budget set in treatment T4. At the other extreme, there are 30 violations arising from a comparison of observed behavior in T2 and predicted behavior in T5, so that 19.2% of all donors that actually made a positive donation under the linear matching treatment T2 are predicted to make a donation that violates GARP when faced with the budget set in treatment T5.

To provide a sense of the magnitude of such violations, Column 2 shows the average donation given among violators of GARP. For example, the first row shows that individuals that violate GARP and donate less than €50 in T4, on average, actually donate €48.1. Hence there are a small number of violations of this prediction of revealed preference theory, and the magnitude of the violations are small. In contrast the second row shows that individuals that violate GARP and donate more than €50 in T1, on average, actually donate €67.6. Hence for this test, there are both a relatively large number of violations, and each violation is quantitatively large.[13]

For comparisons involving the leverage treatment T5, Column 3 restricts the sample to high valuation recipients who would likely donate more than €20 even absent any match, to avoid confounding the comparisons with a change in the identity of the marginal donor. When focusing on high valuation donors, the number of violations falls considerably—for example there are no violations of GARP for high valuation donors from the comparison of treatments T4 and T5. This highlights that some of the earlier violations are likely driven by changes in the composition of donors across treatments. In particular there are likely to be low valuation donors that give positive amounts in the leverage treatment T5 but that would not have donated in any other counterfactual treatment. To take this point one step further, we find a significantly higher proportion of violations if we repeat our method but form predicted donations for *all* mail-out recipients—even those that do not contribute anything in their assigned donation. This is as expected because the observable characteristics of non-donors differ significantly from those of donors, as shown in Table A3.

In summary, from each pairwise comparison of treatments, at least 80% of recipients' behavior is therefore consistent with GARP, so that we can infer that a continuous, convex, and monotone

[12] Gross [1995] provides a detailed discussion of the various measures of goodness of fit that can be employed.

[13] The results are robust to allowing for heterogenous effects in each treatment so that the predicted donation is based on a regression in which each characteristic is also interacted with the relevant treatment dummy.

utility function could have generated their choices.

Whether 80% is considered a large or small number depends on the power of our test, which in turn requires a specific alternative hypothesis to be specified [Varian 1982, Bronars 1995]. On the one hand, in contrast to non-experimental methods, our field experiment allows us to engineer large changes in relative prices holding everything else equal. This improves the power of our test. On the other hand, the bundle at which the budget sets intersect in any two treatments in our design is distant from the bundle chosen on average in the treatments, thus lowering the power of our test. The extent to which these factors offset one another varies across each of the pairwise comparisons in Table 5.

To provide a sense of which of the pairwise comparisons therefore are most informative, we consider the following alternative hypothesis. We generate predicted choices of each donor by first estimating a specification analogous to (4) and impose the restriction that $\beta_3 = 0$, so that under this specific alternative it is assumed that individuals ignore the budget set they face and their donation is driven purely by the observable characteristics listed in Table A1. Column 4 of Table 5 then shows the number and percentage of violations of GARP that would have occurred under this particular alternative hypothesis.

For six out of the ten pairwise comparisons—those based on a comparison of T2 and T4, T2 and T5, T3 and T5—the number of violations based on this alternative are always at least as large as the actual number of violations. In some cases, the number of violations is orders of magnitude smaller than would be expected from this alternative hypothesis, suggesting these pairwise comparisons are powerful tests of GARP. For example, in the comparison between observed donations in T4 and predicted donations in T2, the actual number of violations is 13 while 73 violations are predicted under the alternative hypothesis. Similarly, comparing observed donations in T5 and predicted donations in T2, the actual number of violations is 2 while 30 violations are predicted under the alternative hypothesis.

In contrast, a few of the other comparisons distinctly lack power against this specific alternative. For example, the comparison of actual behavior in T5 to predicted behavior in T3 yields zero violations of GARP under both our test and this alternative hypothesis, so this particular comparison is not informative of whether individual behavior can be rationalized in this setting.

These findings highlight that although the methodological approach of using a field experiment to test for GARP has many advantages over laboratory or non-experimental approaches, the mere fact that large price changes can be induced is not sufficient to guarantee that tests of GARP have high power against an alternative hypothesis. In future work, more powerful tests could be implemented by ensuring the budget lines intersect at bundles closer to observed behavior of more individuals.[14]

[14] We note that for two of the ten pairwise comparisons—between actual donations in T4 and predicted donations in T1, and actual donations in T4 and predicted donations in T5—the actual violations is higher than would be predicted by this specific alternative. In the second of these cases the actual number of violations falls to zero when we consider high valuation donors, as shown in Column 3, although in the first case, the data suggests a different alternative hypothesis should be considered.

4.4 Focal Point Effects?

The results in Table 5 highlight that the pairwise comparisons that yield the highest percentage of violations all involve the non-convex treatment T4. As discussed earlier, this is likely due to the fact that individuals that would otherwise have given less than €50 in T1 choose to donate slightly more than would be predicted by revealed preference theory, that implies they should locate at the interior corner solution and donate precisely €50 in T4.

In addition, the wording in which treatment T4 is described in the mail out letter might lead to d_g = €50 becoming focal in the minds of recipients. If so, then relative to T3 there ought to be bunching in the distribution of donations given *from above* at d_g = €50 in T4. No such bunching above d_g = €50 is predicted in the standard model of consumer choice—this segment of the budget line is available under both T3 and T4. To explore this hypothesis further, we use quantile regression methods to characterize the effect of being assigned to treatment T_i on different percentiles of the conditional distribution of donations given, d_g. This allows us to estimate changes in the shape and spread of the conditional distribution of donations given, not just the change in the mean as estimated in (4). We therefore estimate the following quantile regression specification at each quantile $\theta \in [0,1]$,

$$Quant_{\theta}(\log(d_{gi})|.) = \beta_{\theta}T_i + \gamma_{\theta}X_i \text{ for } d_{gi} > 0. \quad (5)$$

The parameter of interest, β_{θ}, measures the difference at the θth conditional quantile of log donations received between the treatment group T4 rather than T3. Figure 2B graphs estimates of β_{θ} from (5) and the associated 95% confidence interval at each quantile when the comparison treatment is T3. We also show the quantile that corresponds to donations given of €50 (€60) across T3 and T4.

The figure shows the distribution of donations given becomes less dispersed in T4 and in particular, this is because there is a bunching of donations given just above €50 in T4 relative to T3. The estimates show that there are recipients that would otherwise have given less than €50 in T3 are those that shift their donations towards €50 and slightly above. There is no evidence that recipients who would have otherwise given much above €50 significantly reduced their donations towards €50 in T4. Indeed, the distribution of donation given is little changed above €70 in line with there being no or very weak focal point effects introduced by the non-convex treatment.[15]

[15] The analysis highlights that by removing a portion of the budget set for donations given less than €50 in T4 relative to T3, most small donors optimally move to the interior corner solution rather than the exterior solution, while large donors are unaffected—the response rates are almost identical in treatments T3 and T4. This is in contrast to some findings in the psychology literature, where consumers are sometimes observed forgoing a decision altogether in the presence of an expanded choice set [Iyengar and Lepper 2000].

5 Conclusions

We have presented evidence from the first large-scale natural field experiment designed to shed light on whether consumer behavior is consistent with the predictions of revealed preference theory. We do so in the context of a field experiment on charitable giving which allows us to vary budget sets experimentally in a straightforward and very natural manner. We find that consumer behavior, on both the extensive and intensive margins of charitable giving, can be rationalized within a standard model of consumer choice in which individuals have preferences over their own consumption and their contribution towards the charitable project. The behavior of at least 80% of recipients is in line with them adhering to GARP, so we infer that there exists a non-satiated, continuous, monotone, and concave utility function, $u(\mathbf{x})$ that rationalizes the data [Afriat 1967]. The data does not support consumer behavior being driven by non-standard behaviors such as focal point influences. In short, in a real world static environment where participants make simple decisions they are familiar with, the predictions of microeconomic theory work well in explaining the observed choices of individuals.

Given the observed prices and chosen bundles, we have focused our analysis on the question of consistency [Varian 2006], namely whether observed behavior is consistent with utility maximization. We have answered the question using an alternative methodological to the existing literature—a natural field experiment. This approach is relatively straightforward to implement, requires the collection of readily available data, allows us to make inference on a large population in a natural economic environment.

Of course, as with non-experimental methods, it is then possible to take our observed data on prices and bundles to – (i) understand when observed behavior is consistent with maximizing a utility function of particular form; (ii) recover the set of utility functions that are consistent with a given set of choices; (iii) forecast what demand would be if consumers faced a new budget set. The literature relating to each of these questions is discussed in detail in Varian [2006]. We do not focus on these issues here because the method for doing so would be similar in experimental and non-experimental data—the key advantage of our approach is to provide a simple mean by which to understand whether consumer behavior is rationalizable in the first place.

Two points are of relevance for the external validity of our findings. First, in our setting the choice problem has neither elements of intertemporal choice, interactions with others, nor uncertainty. Indeed, we would expect that adding such layers of complexity to the individual choice problem would provide a more significant role for theories in behavioral psychology to explain microeconomic behavior [Armstrong and Spiegler 2007, Della Vigna 2007].

Second, the sample of individuals whose behavior we study—opera attendees—are likely to have higher levels of income and education than the average citizen in the population. It remains an open question as to whether the neoclassical model performs equally well when considering the behaviors of individuals in other income and education classes. Recent work by Burks *et al* [2007] and Huck and Muller [2007] indeed suggests that, for example, IQ and education are correlated to whether subjects display rational behavior in the laboratory and field.

We view both variants on our design—in terms of the complexity of the choice problem and the target population whose decisions are analyzed—as important and exciting areas for the design of future field experiments. We have begun to explore the first of these issues in a follow up experiment with the same opera house [Huck and Rasul 2008]. More precisely, a month after the original mail outs, recipients that did not respond were sent a reminder letter. This stated that the original matching scheme, if any, was no longer in place, but still encouraged individuals to donate to the project. Some of the results of this experiment were discussed in Section 4 to shed light on the nature of selection into response, but this follow-up experiment can also be used to understand whether the history of prices consumers face affects their behavior in any given period. Preliminary results suggest this is the case, and may, in part, be driven by the regret individuals feel at not having previously contributed to the charitable good when its price was relatively low. Although preliminary, these findings suggest adding a small layer of complexity—introducing dynamics—leads the choice data to be best explained by a combination of neoclassical theory with elements of behavioral economics.

References

[1] AFRIAT.S (1967) "The Construction of Utility Functions from Expenditure Date", *International Economic Review* 8: 67-77.

[2] ANDREONI.J (1990) "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving", *Economic Journal* 100: 464-77.

[3] ANDREONI.J (1998) "Toward a Theory of Charitable Fund-Raising", *Journal of Political Economy* 106: 1186-213.

[4] ANDREONI.J (2006a) "Leadership Giving in Charitable Fund-Raising", *Journal of Public Economic Theory* 8: 1-22.

[5] ANDREONI.J (2006b) "Philanthropy", in *The Handbook of Giving, Reciprocity and Altruism*, S.C.Kolm and J.Mercier Ythier (eds.), Amsterdam: North Holland.

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

[7] ANGRIST.J.D (1997) "Conditional Independence in Sample Selection Models", *Economics Letters* 54: 103-12.

[8] ARMSTRONG.M AND R.SPIEGLER (2007), Boundedly Rational Consumers: Outline of a Survey, mimeo UCL.

[9] BATTALIO.R.C, J.H.KAGEL, R.C.WINKLER, E.B.FISHER, R.L.BASMANN, AND L.KRANSER (1973) "A Test of Consumer Demand Theory Using Observations on Individual Consumer Purchases", *Western Economic Journal* 11: 411-28.

[10] BLUNDELL.R, M.BROWNING, AND I.CRAWFORD (2003) "Nonparametric Engel Curves and Revealed Preference", *Econometrica* 71: 205-40.

[11] BLUNDELL.R, M.BROWNING, AND I.CRAWFORD (2007) Best Nonparametric Bounds on Demand Responses, Cemmap Working Paper 12/05.

[12] BRONARS.S.G (1995) "The Power of Nonparametric Tests of Preference Maximisation", *Econometrica* 55: 693-98.

[13] BURKS.S.V, J.CARPENTER, L.GOTTE, K.MONACO, K.PORTER, AND A.RUSTICHINI (2007) Using Behavioral Economic Experiments at a Large Motor Carrier: The Context and Design of the Truckers and Turnover Project, IZA Discussion Paper 2789.

[14] CHAY.K.Y AND J.POWELL (2001) "Semiparametric Censored Regression Models", *Journal of Economic Perspectives* 15: 29-42.

[15] CHOI.S, R.FISMAN, D.GALE, AND S.KARIV (2007) "Consistency and Heterogeneity of Individual Behavior under Uncertainty", forthcoming *American Economic Review*.

[16] COX.J.C (1997) "On Testing the Utility Hypothesis", *Economic Journal* 107: 1054-78.

[17] DELLA VIGNA.S (2007) "Psychology and Economics: Evidence from the Field", forthcoming, *Journal of Economic Literature*.

[18] DOHMEN.T, A.FALK, D.HUFFMAN, AND U.SUNDE (2006) Seemingly Irrelevant Events Affect Economic Perceptions and Expectations: The FIFA World Cup 2006 as a Natural Experiment, IZA Discussion Paper 2275.

[19] ECKEL.C AND P.GROSSMAN (2006) Subsidizing Charitable Giving: A Field Test Comparing Matching and Rebate Subsidies, mimeo, Virginia Tech.

[20] FAMULARI.M (1995) "A Household-Based, Nonparametric Test of Demand Theory", *Econometrica* 77: 372-83.

[21] GROSS.J (1995) "Testing Data for Consistency With Revealed Preference", *Review of Economics and Statistics* 77: 701-10.

[22] 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.

[23] HARDLE.W, W.HILDENBRAND, AND M.JERISON (1991) "Empirical Evidence on the Law of Demand", *Econometrica* 59: 1525-49.

[24] HARRISON.G AND J.A.LIST (2004) "Field Experiments", *Journal of Economic Literature* 42: 1009-55.

[25] HECKMAN.J.J, LALONDE.R, AND J.SMITH (1999): "The Economics and Econometrics of Active Labor Market Programs" in *Handbook of Labor Economics*, Volume III, O.Ashenfelter and D.Card (editors), Elsevier: Amsterdam.

[26] HOUTHAKKER.H.S (1950) "Revealed Preference and the Utility Function", *Economica* 17: 159-74.

[27] HUCK.S AND W.MULLER (2007) Allais For All: Revisiting the Paradox, mimeo, University College London.

[28] HUCK.S AND I.RASUL (2007) Comparing Charitable Fundraising Schemes: Evidence From a Natural Field Experiment, mimeo, University College London.

[29] HUCK.S AND I.RASUL (2008) The Economics of Reminders: Evidence From a Natural Field Experiment, mimeo, University College London.

[30] IYENGAR.S AND M.LEPPER (2000) "When Choice is Demotivating: Can One Have Too Much of a Good Thing?", *Journal of Personality and Social Psychology* 79: 995-1006.

[31] KARLAN.D AND J.A.LIST (2007) "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment", forthcoming *American Economic Review.*

[32] KOO.A.Y.C (1962) "An Empirical Test of Revealed Preference Theory", *Econometrica* 31: 646-64.

[33] LIST.J.A AND D.LUCKING-REILEY (2002) "The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign", *Journal of Political Economy* 110: 215-33.

[34] LIST.J.A AND D.MILLIMET (2005) Bounding the Impact of Market Experience on Rationality: Evidence from a Field Experiment with Imperfect Compliance, mimeo University of Chicago.

[35] MANSER.M.E AND R.J.MCDONALD (1988) "An Analysis of Substitution Bias in Measuring Inflation, 1959-85", *Econometrica* 56: 909-30.

[36] MOSSIN.A (1972) "A Mean Demand Function and Individual Demand Functions Confronted with the Weak and the Strong Axioms of Revealed Preference: An Empirical Test", *Econometrica* 40: 177-92.

[37] RONDEAU.D AND J.A.LIST (2008) "Matching and Challenge Gifts to Charity: Evidence from Laboratory and Natural Field Experiments", *Experimental Economics* 11: 253-81.

[38] ROSE.H (1958) "Consistency of Preference: The Two Commodity Case", *Review of Economic Studies* 25: 124-5.

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

[40] SIPPEL.R (1997) "An Experiment on the Pure Theory of Consumer's Behaviour", *Economic Journal* 107: 1431-44.

[41] VARIAN.H.R (1982) "The Nonparametric Approach to Demand Analysis", *Econometrica* 50: 945-74.

[42] VARIAN.H.R (2006) "Revealed Preference", in M.Szenberg (editor), *Samuelsonian Economics and the 21st Century*, Oxford University Press, Oxford, England.

[43] WHITNEY.G.A AND J.L.SWOFFORD (1987) "Nonparametric Tests of Utility Maximization and Weak Separability for Consumption, Leisure, and Money", *Review of Economics and Statistics* 69: 458-64.

[44] WOOLDRIDGE.J.M (2002) Econometric Analysis of Cross Section and Panel Data, Cambridge, Mass.: MIT Press.

Table 1: Outcomes by Treatment—Descriptive Evidence

Mean, standard error in parentheses

P-values on tests of equalities on means with comparison group in brackets

Treatment Number	Treatment Description	Comparison Group	Total Amount Donated (1)	Total Amount Raised (2)	Number of Donors (3)	Response Rate (4)	Average Donation Received (5)	Median Donation Received (6)	Average Donation Given (7)	Median Donation Given (8)
1	**Lead donor (Control)**		17416	17416	132	.035	132	100	132	100
						(.003)	(14.3)		(14.3)	
2	**Lead donor + 1:.5 matching**		15705	23558	156	.042	151	75	101	50
						(.003)	(18.9)		(12.6)	
		T1 Control				[.134]	[.421]	[.131]	[.102]	[.000]
3	**Lead donor + 1:1 matching**		14310	28620	155	.042	185	100	92.3	50
						(.003)	(20.7)		(10.4)	
		T1 Control				[.133]	[.037]	[.999]	[.025]	[.000]
4	**Lead donor + 1:1 matching for donations greater than €50**		15671	31107	160	.043	194	120	97.9	60
						(.003)	(19.3)		(9.59)	
		T1 Control				[.084]	[.010]	[.102]	[.049]	[.000]
5	**Lead donor + €20 match for any donation**		12248	15788	177	.047	89.2	70	69.2	50
						(.003)	(5.51)		(5.51)	
		T1 Control				[.008]	[.006]	[.065]	[.000]	[.002]

Notes: The test of equality of means is based on an OLS regression allowing for robust standard errors. The test of equality of medians is based on a quantile regression. The total amount raised corresponds to the sum of donations of all individual recipient observations. The response rate is the proportion of recipients that donate some positive amount, as reported in the donation amount column. The actual donation

Table 2: Pairwise Tests of Revealed Preference

P-value on relevant test for each column is in brackets

Treatments Being Compared		Type of Comparison	Response Rate [One Sided t-test]	Proportions Above/Below Some Critical Value [One Sided t-test]		Distribution of Donations Given [Mann Whitney test]
T1: Lead donor (control)	**T2: Lead donor + 1:.5 match**	Budget set expands	Weakly higher in T2 [.067]			
T1: Lead donor (control)	**T3: Lead donor + 1:1 match**	Budget set expands	Weakly higher in T3 [.066]			
T1: Lead donor (control)	**T4: Lead donor + 1:1 match for donations greater than €50**	Budget set expands and partly coincides	Weakly higher in T4 [.042]			
T2: Lead donor + 1:.5 match	**T4: Lead donor + 1:1 match for donations greater than €50**	Budget sets cross		Proportion of donations < 50 weakly higher in T2	Proportion of donations > 50 weakly higher in T4 [.000]	
T2: Lead donor + 1:.5 match	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Weakly higher in T5 [.120]	Proportion of donations < 40 weakly higher in T5	Proportion of donations > 40 weakly higher in T2 [.014]	
T3: Lead donor + 1:1 match	**T4: Lead donor + 1:1 match for donations greater than €50**	Budget set expands and partly coincides	Weakly higher in T3 [.587]			Identical for donations > 50 (if no focal point effects) [.000]
T3: Lead donor + 1:1 match	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Weakly higher in T5 [.122]	Proportion of donations < 20 weakly higher in T3	Proportion of donations > 20 weakly higher in T5 [.013]	
T4: Lead donor + 1:1 match for donations greater than €50	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Weakly higher in T5 [.172]	Proportion of donations < 50 weakly higher in T5	Proportion of donations > 50 weakly higher in T4 [.000]	

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Tests that are outlined are those that require the assumption of convexity on consumer preferences. The tests of proportions are based on all mail out recipients.

Table 3: Outcomes by Treatment—Extensive Margin

Marginal effects reported in probit regressions

Robust standard errors in parentheses

Dependent variable:	**Match Response (Probit Regressions)**				
Comparison treatment:	**T1**	**T1**	**T3**	**T1**	**T1: Predicted High Donors**
	(1)	**(2a)**	**(2b)**	**(3a)**	**(3b)**
50% Match Treatment T2	.007				
	(.005)				
100% Match Treatment T3	.007				
	(.005)				
Non-convex Treatment T4		.008*	.001		
		(.004)	(.004)		
Leveraged Treatment T5				.013***	.010
				(.004)	(.007)
Observations	11233	7516	7464	7516	3317

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Robust standard errors estimated throughout. The dependent variable is equal to one if the recipient responds to the matching treatment with any positive donation, and zero otherwise. Marginal effects are reported. The sample in Column 3b is restricted to those recipients that are predicted to donate higher than average amounts (absent any match). All specifications control for the recipient's gender, the number of ticket orders placed in the 12 months prior to mail out, the average price of these tickets, whether the recipient is a Munich resident, and a dummy variable for whether the year of the last ticket purchase was 2006 or not.

Table 4: Outcomes by Treatment—Intensive Margin (OLS and Hurdle Model Estimates)

Robust standard errors in parentheses

Dependent variable:	Donation Received (in levels for OLS estimates, in logs for hurdle model estimates)											
Comparison treatment:	**T1**	**T1**	**T1**	**T1**	**T3**	**T3**	**T3: Predicted Very High Donors**	**T3: Predicted Very High Donors**	**T1**	**T1**	**T1: Predicted High Donors**	**T1: Predicted High Donors**
	(1a) OLS	**(1b) Second Stage Hurdle Model**	**(2a) OLS**	**(2b) Second Stage Hurdle Model**	**(2c) OLS**	**(2d) Second Stage Hurdle Model**	**(2e) OLS**	**(2f)**	**(3a) OLS**	**(3b) Second Stage Hurdle Model**	**(3c) OLS**	**(3d) Second Stage Hurdle Model**
50% Match Treatment T2	1.61	.178										
	(1.12)	(.128)										
100% Match Treatment T3	3.07**	.457***										
	(1.22)	(.124)										
Non-convex Treatment T4			3.80***	.532***	.795	.180**	1.82	.044				
			(1.22)	(.095)	(1.45)	(.085)	(4.62)	(.165)				
Leveraged Treatment T5									-.372	.012	-1.43	-.193
									(.750)	(.112)	(1.49)	(.154)
Observations	11233	443	7516	292	7464	315	1860	101	7516	309	3317	160

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Robust standard errors estimated throughout. In Columns 1a, 2a, 2c, 2e, 3a, and 3c an OLS regression is run where the dependent variable is the amount received by the opera house (including any match) from the recipient. In Columns 1b, 2b, 2d, 2f, 3b, and 3d the second stage of a hurdle model is estimated assuming the donation amounts follow a log normal distribution. The dependent variable is the log of the donation received. The samples in Columns 2e and 2f are restricted to those recipients that are predicted to be in the top quantile of donation amounts (absent any match). The samples in Columns 3c and 3d are restricted to those recipients that are predicted to give higher than average amounts (absent any match). All specifications control for the recipient's gender, the number of ticket orders placed in

Table 5: Individual Violations of Revealed Preference

Matching Treatments Being Compared		Type of Comparison	GARP Violation	Number (Percentage) of Violations (1)	Donation Given Among Violators [95% confidence interval] (2)	Number (Percentage) of Violations, Predicted High Donors (3)	Alternative Hypothesis: Number (Percentage) of Violations (4)
T1: Lead donor (control)	**T4: Lead donor + 1:1 match for donations greater than €50**	Budget set expands and partly coincides	Give more than 50 in T1 and predicted to give less than 50 in T4	3 (2.27)	48.1 [43.0, 53.1]		6 (4.45)
			Give less than 50 in T4 and predicted to give more than 50 in T1	27 (16.9)	67.6 [57.7, 77.6]		16 (10.0)
T2: Lead donor + 1:.5 match	**T4: Lead donor + 1:1 match for donations greater than €50**	Budget sets cross	Give more than 50 in T2 and predicted to give less than 50 in T4	3 (1.92)	48.1 [43.0, 53.1]		5 (3.21)
			Give more than 50 in T4 and predicted to give less than 50 in T2	13 (8.13)	43.8 [40.9, 46.6]		73 (45.6)
T2: Lead donor + 1:.5 match	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Give less than 40 in T2 and predicted to give more than 40 in T5	30 (19.2)	78.6 [68.5, 88.8]	9 (7.6)	50 (32.1)
			Give more than 40 in T5 and predicted to give less than 40 in T2	2 (1.12)	28.5 [-2.66, 59.7]	0 (0)	30 (16.9)
T3: Lead donor + 1:1 match	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Give less than 20 in T3 and predicted to give more than 20 in T5	5 (3.23)	61.0 [33.9, 88.1]	-	20 (12.9)
			Give more than 20 in T5 and predicted to give less than 20 in T3	0 (0)	-	-	0 (0)
T4: Lead donor + 1:1 match for donations greater than €50	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Give less than 50 in T4 and predicted to give more than 50 in T5	27 (16.9)	80.6 [70.2, 91.0]	0 (0)	10 (6.25)
			Give more than 50 in T5 and predicted to give less than 50 in T4	3 (1.69)	48.1 [43.0, 53.1]	0 (0)	5 (2.82)

Notes: The number of violations are based on recipients that responded with some positive donation in their assigned treatment. In Columns 1 and 4 the proportion of violations is the number of violations divided by the total number of positive donations given in the treatment from which actual (and not predicted) donations are used. Column 2 shows the predicted donation in each pairwise comparison among those individuals that violate the predictions of revealed preference theory. The pairs in Column 3 are restricted to those that are predicted to give higher than average amounts (absent any match). In Column 4 we form predicted donations by regressing the log of donations received on observable characteristics of the recipient but not the treatment dummy.

Table A1: Characteristics of Recipients by Matching Treatment

Mean, standard error in parentheses
P-value on test of equality of means with control group in brackets

Treatment Number	Treatment Description	Number of Individuals	Female [Yes=1]	Number of Tickets Bought in Last 12 Months	Number of Ticket Orders in Last 12 Months	Average Price of Tickets Bought in Last 12 Months	Total Value of All Tickets Bought in Last 12 Months	Munich Resident [Yes=1]	Year of Last Ticket Purchase [2006=1]
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	**Lead donor (Control)**	3770	.478	6.27	2.22	86.3	423	.416	.574
			(.008)	(.153)	(.046)	(.650)	(7.73)	(.008)	(.008)
2	**Lead donor + 1:.5 match**	3745	.481	6.39	2.20	86.8	432	.416	.576
			(.008)	(.184)	(.049)	(.660)	(9.63)	(.008)	(.008)
			[.818]	[.606]	[.851]	[.603]	[.451]	[.989]	[.863]
3	**Lead donor + 1:1 match**	3718	.477	6.46	2.28	85.8	435	.419	.576
			(.008)	(.148)	(.050)	(.667)	(9.78)	(.008)	(.008)
			[.923]	[.362]	[.329]	[.642]	[.314]	[.838]	[.890]
4	**Lead donor + 1:1 match for donations greater than €50**	3746	.476	6.31	2.21	85.2	419	.426	.567
			(.008)	(.145)	(.046)	(.657)	(7.39)	(.008)	(.008)
			[.825]	[.832]	[.949]	[.238]	[.726]	[.399]	[.540]
5	**Lead donor + €20 match for any donation**	3746	.486	6.09	2.20	86.5	416	.428	.556
			(.008)	(.132)	(.047)	(.657)	(8.05)	(.008)	(.008)
			[.525]	[.404]	[.765]	[.855]	[.578]	[.281]	[.108]

Notes: The tests of equality are based on an OLS regression allowing for robust standard errors. All monetary amounts are measured in Euros. The "last twelve months" refers to the year prior to the mail out from June 2005 to June 2006.

Table A2: Characteristics of Donors by Treatments

P-value on test of equality of MEANS with control treatment T1 in parentheses
P-value on test of equality of DISTRIBUTION with control treatment T1 in brackets

Treatment Number and Description		Comparison Group T1: Lead Donor (control)	Female [Yes=1]	Number of Tickets Bought in Last 12 Months	Number of Ticket Orders in Last 12 Months	Average Price of Tickets Bought in Last 12 Months	Total Value of All Tickets Bought in Last 12 Months	Munich Resident [Yes=1]	Year of Last Ticket Purchase [2006=1]
2	Lead donor + 1:.5 match	Equality of means	(.169)	(.913)	(.279)	(.886)	(.958)	(.620)	(.102)
		Equality of distributions	-	[.859]	[.701]	[.732]	[.856]	-	-
3	Lead donor + 1:1 match	Equality of means	(.078)	(.200)	(.186)	(.896)	(.592)	(.745)	(.908)
		Equality of distributions	-	[.161]	[.190]	[.684]	[.109]	-	-
4	Lead donor + 1:1 match for donations greater than €50	Equality of means	(.584)	(.235)	(.066)	(.622)	(.741)	(.165)	(.358)
		Equality of distributions	-	[.289]	[.430]	[.257]	[.413]	-	-
5	Lead donor + €20 match for any donation	Equality of means	(.441)	(.432)	(.516)	(.409)	(.635)	(.131)	(.258)
		Equality of distributions	-	[.332]	[.821]	[.759]	[.172]	-	-

Notes: The tests of equality of means are based on an OLS regression allowing for robust standard errors. The test of the equality of distributions is based on the Mann Witney test statistic. The null hypothesis is that the two independent samples are drawn from populations with the same distribution. We perform this test only for continuous variables and not the dummy variables. All monetary amounts are measured in Euros. The "last twelve months" refers to the year prior to the mail out from June 2005 to June 2006.

Table A3: Characteristics of Donors and Non Donors by Treatment

Mean, standard error in parentheses

P-value on test of equality of means with comparison group in brackets

Treatment Number and Description	Comparison Group	Number (Proportion) of Responses	Response Rate	Female [Yes=1]	Number of Tickets Bought in Last 12 Months	Number of Ticket Orders in Last 12 Months	Average Price of Tickets Bought in Last 12 Months	Total Value of All Tickets Bought in Last 12 Months	Munich Resident [Yes=1]	Year of Last Ticket Purchase [2006=1]
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 Lead donor (control)		132	.035	.530	9.52	3.36	88.1	666	.394	.742
		(14.3)	(.003)	(.044)	(.954)	(.374)	(4.21)	(66.1)	(.043)	(.038)
	Non Donors in Same Treatment			[.226]	[.000]	[.002]	[.669]	[.000]	[.600]	[.000]
2 Lead donor + 1:.5 match		156	.042	.449	9.67	2.88	87.3	662	.365	.654
		(16.9)	(.003)	(.040)	(1.09)	(.235)	(3.33)	(55.1)	(.039)	(.038)
	Non Donors in Same Treatment			[.412]	[.002]	[.004]	[.874]	[.000]	[.183]	[.040]
3 Lead donor + 1:1 match		155	.042	.426	7.96	2.75	88.9	619	.413	.748
		(16.8)	(.003)	(.040)	(.746)	(.274)	(4.13)	(58.2)	(.040)	(.035)
	Non Donors in Same Treatment			[.190]	[.041]	[.083]	[.452]	[.001]	[.886]	[.000]
4 Lead donor + 1:1 match for donations greater than €50		160	.043	.563	8.11	2.60	90.7	698	.475	.694
		(17.4)	(.003)	(.039)	(.703)	(.177)	(3.14)	(69.7)	(.040)	(.037)
	Non Donors in Same Treatment			[.025]	[.010]	[.028]	[.079]	[.000]	[.206]	[.000]
5 Lead donor + €20 match for any donation		177	.047	.486	8.58	3.07	83.8	623	.480	.684
		(19.2)	(.003)	(.038)	(.719)	(.243)	(2.98)	(63.9)	(.038)	(.035)
	Non Donors in Same Treatment			[.994]	[.000]	[.000]	[.367]	[.001]	[.160]	[.003]

Notes: All figures refer to the mail out recipients in each treatment excluding non-German residents, corporate donors, formally titled donors, and recipients to whom no gender can be assigned. The tests of equality are based on an OLS regression allowing for robust standard errors. All monetary amounts are measured in Euros. In Column 1, the proportion refers to the proportion of all donors that were in the given treatment. In Columns 4 to 7 the "last twelve months" refers to the year prior to the mail out from June 2005 to June 2006.

Figure 1: The Design of the Field Experiment and Outcomes by Treatment

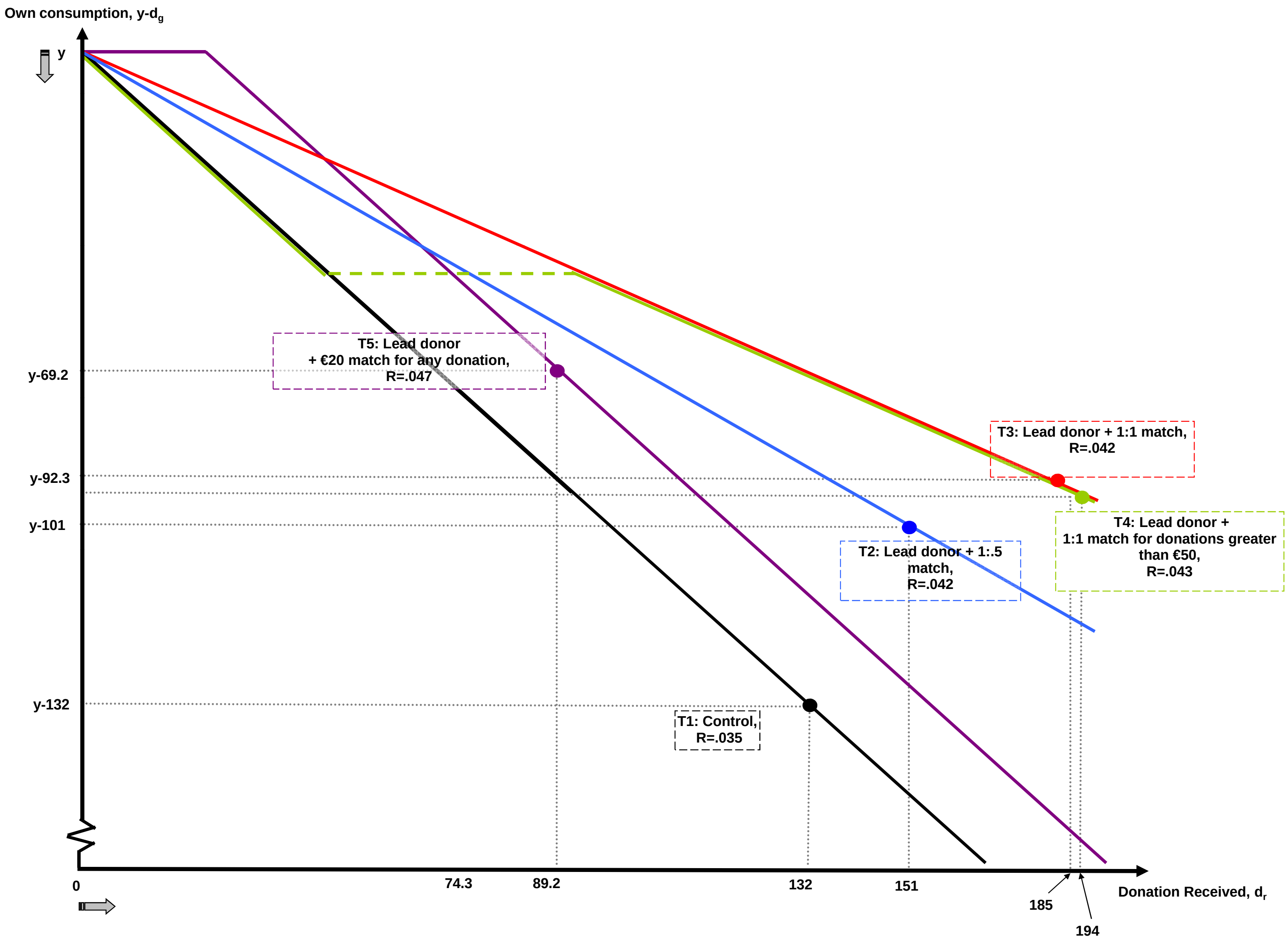


Figure 2A: Proportion of Donations Given, T4 minus T3

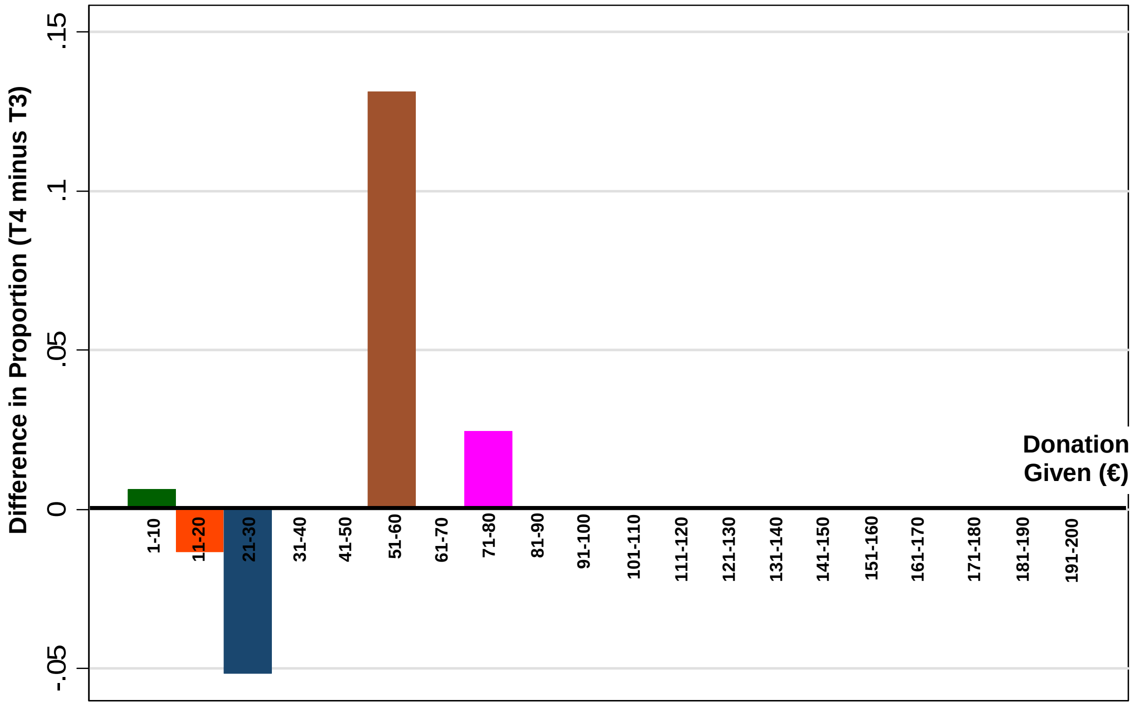
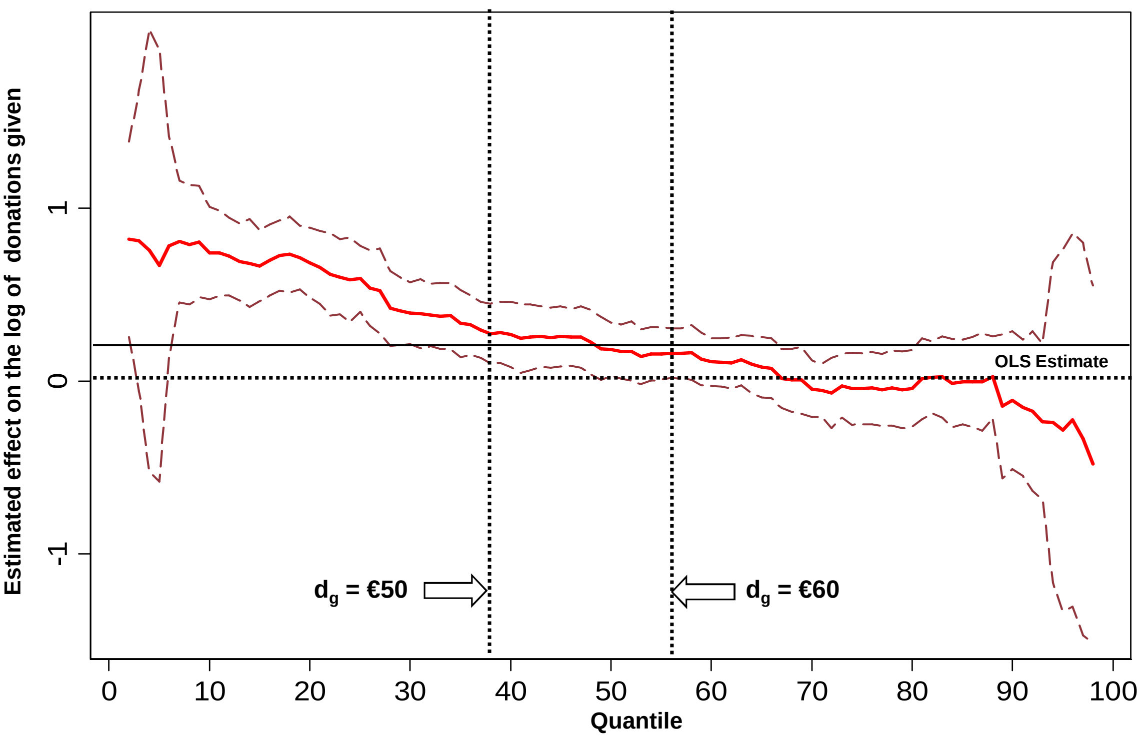


Figure 2B: Non Convex Budget Set T4 Relative to 100% Match Rate Treatment T3



Notes: Figure 2A plots the difference in the proportion of donations given by 10 Euro bins. A positive value indicates that proportionally more donations were given in the bin under T4 than T3, and negative values have the opposite interpretation. Figure 2B shows the estimated effect of being assigned to the non-convex treatment T4 relative to being assigned to the 100% matching treatment T3 on the log of donations given, at each quantile of the conditional distribution of the log of donations given, and the associated 95% confidence interval. The figure also shows the coefficient on the treatment dummy variable from an OLS regression. The individual characteristics controlled for are whether the recipient is female, the number of ticket orders placed in the 12 months prior to mail out, the average price of these tickets, whether the recipient is a Munich resident, and a dummy variable for whether the year of the last ticket purchase was 2006 or not.

Appendix: The Mail Out Letter (Translated)

Bayerische Staatsoper
Staatsintendant
Max-Joseph-Platz 2, D-80539 München
www.staatsoper.de

[ADDRESS OF RECIPIENT]

Dear [RECIPIENT],

The Bavarian State Opera House has been investing in the musical education of children and youths for several years now as the operatic the art form is in increasing danger of disappearing from the cultural memory of future generations.

Enthusiasm for music and opera is awakened in many different ways in our children and youth programme, "Erlebnis Oper" *[Experience Opera]*. In the forthcoming season 2006/7 we will enlarge the scope of this programme through a new project "Stück für Stück" that specifically invites children from schools in socially disadvantaged areas to a playful introduction into the world of opera. Since we have extremely limited own funds for this project, the school children will only be able to experience the value of opera with the help of private donations.

[This paragraph describes each matching scheme and is experimentally varied as described in the main text of the paper].

As a thank you we will give away a pair of opera tickets for Engelbert Humperdinck's "Konigskinder" on Wednesday, 12 July 2006 in the music director's box as well as fifty CDs signed by Maestro Zubin Mehta among all donors.

You can find all further information in the enclosed material. In case of any questions please give our Development team a ring on *[phone number]*. I would be very pleased if we could enable the project "Stück für Stück" through this appeal and, thus, make sure that the operatic experience is preserved for younger generations.

With many thanks for your support and best wishes,

Sir Peter Jonas, Staatsintendant

Appendix: The Mail Out Letter (Translated)

“Stück für Stück”

The project “Stück für Stück” has been developed specifically for school children from socially disadvantaged areas. Musical education serves many different functions in particular for children and youths with difficult backgrounds -- it strengthens social competence and own personality, improves children’s willingness to perform, and reduces social inequality. Since music education plays a lesser and lesser role in home and school education, the Bavarian State Opera has taken it on to contribute to it ourselves. The world of opera as a place of fascination is made attainable and accessible for young people.

In drama and music workshops, “Stück für Stück” will give insights into the world of opera for groups of around 30 children. They will be intensively and creatively prepared for a subsequent visit of an opera performance. These workshops encourage sensual perception – through ear and eye but also through scenic and physical play and intellectual comprehension – all of these are important elements for the workshops. How does Orpheus in “Orphee and Eurydice” manage to persuade the gods to let him save his wife from the realm of dead? Why does he fail? Why poses the opera “Cosi fan tutte” that girls can never be faithful? It is questions like these that are investigated on the workshops.

The workshops are also made special through the large number and variety of people who are involved in them: musicians, singers, directors, and people from many other departments, ranging from costumes and makeup to marketing. The participants in each workshop work through an opera’s storyline, and are introduced to the production and will meet singers in their costumes as well as musicians. This makes the workshops authentic. After the workshops the participants are invited to see the actual opera production.

Through your donation the project “Stück für Stück” will be made financially viable so that we can charge only a small symbolic fee to the participants. This makes it possible to offer our children and youth programme also to children from socially disadvantaged backgrounds that can, thus, learn about the fascination of opera.

Note: In German, Stück für Stück is a wordplay --- “Stück” meaning “play” as in drama and “Stück für Stück” being an expression for doing something bit by bit.