RATIONALITY
& COMPETITION
CRC TRR 190

Testing Consumer Theory: Evidence From a Natural Field Experiment

Maja Adena (WZB)
Steffen Huck (WZB)
Imran Rasul (UCL)

Discussion Paper No. 54

November 2, 2017

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

TESTING CONSUMER THEORY:
EVIDENCE FROM A NATURAL FIELD EXPERIMENT

MAJA ADENA (WZB) STEFFEN HUCK (WZB & UCL)
IMRAN RASUL (UCL)*

SEPTEMBER 2017

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. 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, satisfying the axioms of revealed preference.

Keywords: natural field experiment; revealed preference.

JEL Classification: C93, D01, D12, D64.

*Acknowledgements: We thank the Editor, Robert Slonim, and one anonymous reviewer for helpful comments.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. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 and from the ESRC is gratefully acknowledged. All errors remain our own.

1 Introduction

Neoclassical 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 shedding 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 [e.g. Afriat 1967]. We do this through the lens of a natural field experiment on charitable giving.

By focusing our analysis on the choice between a charitable good and private consumption, we vary the budget set individuals face in a straightforward and natural way, holding all other prices constant. We do so by offering various *matching schemes* that affect how donations given for the charitable good translate into donations received by the project. Specifically, we induce—(i) large changes in the relative price of the charitable good through 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) a non-convex budget set in which only donations above some threshold are matched.

In our design the induced budget sets intersect each other, opening up the possibility to directly test the predictions of revealed preference theory. For such research questions, a between-subject research design is strictly preferred to a within-subject design. This is because 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.

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. 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 highlight that field experiments can be used to test revealed preference theory and such approaches are complementary to non-experimental tests of consumer theory which typically exploit panel data on consumer purchases. However, as in 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 and consumption expenditures are typically measured with error. 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 [Mossin 1972, Hardle *et al* 1991] and others finding more rationalizable patterns of consumption [Manser and McDonald 1988, Famulari 1995]. 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 *et al* 1973, Cox 1997, Sippel 1997, Andreoni and Miller 2002, Choi *et al* 2007, List and Millimet 2008]. Our research design combines the key advantages of laboratory experiments in being able to experimentally manipulate the economic environment faced by agents 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.[1] [2]

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. The project's beneficiaries are children from disadvantaged families whose parents are almost surely not among the recipients of the mail out. 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.

Individuals were randomly assigned to one of five treatments that varied in how individual donations would be matched by an anonymous lead donor. The 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. Since the presence of a lead donor may serve as a signal of project quality [Vesterlund 2003, Andreoni 2006], it is essential that the lead donor is also mentioned in a baseline treatment. Hence in the control treatment T1, recipients were informed that the project had already garnered a lead gift of €60,000, but there was no offer to match donations. The wording of the key paragraph read as follows,

[1] 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.

[2] Our analysis here focuses on the broad question of whether individual behavior is consistent with neoclassical microeconomic theory. In companion papers we exploit the natural field experiment to shed light on specific issues relating to the economics of charitable giving [Huck and Rasul 2011, and Huck, Rasul, and Shephard 2015].

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.*

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

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.*

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.*

T5 (Income): *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.*

Notice how T4 and T5 generate budget constraints that overlap and cross with others thus generating revealed preference predictions.

2.2 Conceptual Framework

We assume potential donors have preferences defined over two dimensions—their own consumption, c, and the marginal benefit their donation provide, 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 indirectly 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 (directly or indirectly) 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. The main result in the revealed preference literature is from Afriat [1967] which states that given some choice data $(\mathbf{p}^t, \mathbf{x}^t)$ for $t = 1, ..., T$, the following conditions are equivalent: (i) the data satisfy GARP; (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 individual behavior being rationalized by 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.

Figure 1 graphs the budget sets induced by the five treatments in $(y - d_g, d_r)$-space. As the budget sets across treatments 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.

3 Descriptives

3.1 Treatment Assignment, Extensive and Intensive Margin Outcomes

Table 1 summarizes information on individuals in each treatment and 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.

Table 2 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, $\mathbf{p}^t$. In our sample of 18,725 individual recipients, Columns 1 to 3 reveal that overall, 780 individuals donated a total of €75,350, corresponding to €116,489 raised for the project, with a mean donation given of €96.6.

On the extensive margin of giving, Column 4 shows that response rates vary from 3.5% to 4.7% across treatments, which are almost double those in comparable large-scale natural field experiments on charitable giving [Eckel and Grossman 2008, Karlan and List 2007]. Indeed, a rule of thumb used by charitable organizations is to expect response rates to mail solicitations of between .5% and 2.5% [de Oliveira *et al* 2011].

On the relative price of giving we note that despite there being large variations in the budget sets in treatments T1 to T3, there are no statistically significant differences in response rates across these treatments. 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. As shown in Figure 1 and Column 7 of Table 2, as the match rate increases, the average donation *given*, d_g, falls from €132 in the control treatment T1 to €101 in T2 with a 50% match rate, and to €92.3 in T3 with a 100% match rate.

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 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.

Treatment T5—that causes a parallel shift out of the budget set conditional on any positive donation, should induce the largest change in the number of donors relative to the control group because *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 response rate is indeed significantly higher in T5 relative to the other treatments. However, it is still only 4.7%, highlighting that even among this targeted population, 95% of individuals do not care for the project. Comparing the income treatment T5 to the control treatment, consumer theory suggests these additional donors should be willing to contribute relatively small amounts to the project which is strongly supported in the data.

4 Testing Revealed Preference Theory

4.1 Aggregate Violations

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: (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 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 3 presents the results for each pairwise treatment comparison. Columns (1)-(3) give the hypotheses to be tested of the type: "the behavior is consistent with revealed preferences." One test 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 theory. Thirteen of the fourteen tests do not reject the hypothesis that consumers, on average, having an underlying utility function that displays standard properties.

The exception is the test between T3 and T4 in the last column that is based on the assumption of convexity. 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 are expected to donate no more than €50 in T4. Hence relative to T3, there ought to be relatively more donations given *below or at* $d_g =$ €50 in T4. In the data there is, however, a bunching of donations in T4 relative to T3 slightly above $d_g =$ €50, and a fall in the proportion of donations given below €50, that is, we find that donors prefer to give incrementally above €50 when faced with the non-convex budget set (perhaps in order to avoid the appearance of being "cheap").

4.2 Individual Violations

In our between-subject design we do not observe the same consumer making multiple choices under alternative budget sets. To detect *individual* violations of GARP we propose a novel approach based on the estimate for each individual i, whose actual choice we only observe in treatment t, for what she would have donated in the relevant counterfactual treatment $t' \neq t$ based on the predictions from a hurdle model. This 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: the choice of d_r conditional on $D_i = 1$. A simple two-tiered model for charitable giving has, as a first stage, a probit model of giving. At the second stage, we assume donations received from individual i are log normally distributed conditional on $d_{ri} > 0$. The maximum likelihood estimator of the second stage parameters is then simply the OLS estimator from the following regression,

$$\log(d_{ri}) = \beta T_i + \gamma X_i + z_i \text{ for } d_{ri} > 0, \tag{2}$$

where T_i is a dummy for any treatment T_i that the individual was assigned to (T2-T5). We estimate the coefficients relative to a control treatment for each treatment separately.[3] We also control for the following individual characteristics X_i, to reduce the sampling errors of the treatment effect estimates: 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. We calculate robust standard errors. More details of the procedure are provided in the Technical Appendix.

In a second step, for each individual and treatment that this individual was not in, we predict her donation amount based on her individual characteristics, fictive treatment assignment and the coefficient estimates from the first stage. We use this comparison between one actual treatment t and one predicted counterfactual treatment t' as the basis of tests for individual violations of revealed preference theory.[4] There are 10 such pairwise comparisons, as Table 4 shows. These are analogous to a subset of the tests performed in Table 3, namely those for which the budget sets intersect. Column 1 shows the number of violations of revealed preference theory for each pairwise comparison of treatments. We also show the proportion of violations defined as the number of violations divided by the number of positive actual donations that fulfill the first part of the condition.[5] Both measures have been previously used in the literature as measures of goodness of fit in tests of revealed preference [Gross 1995].

Across pairwise comparisons, the proportion of violations varies. To provide a sense of the magnitude of such violations, Column 2 shows the average donation given *among* violators of GARP and a 95% confidence interval. The first row shows that individuals that violate GARP and donate less than €50 in T4, on average, actually donate €49.5. Hence there are a small number of violations of this prediction of revealed preference theory, and the magnitude of the violations is small. In contrast, the fifth row shows that individuals that violate GARP and donate more than €40 in T5, on average, actually donate €68. Hence for this test, there are both a relatively large number of violations, and violations are quantitatively large.

For comparisons involving the income treatment T5, Column 3 restricts the sample to high valuation recipients who, based on their predicted donation from (2), would likely donate more

[3] The omitted treatment is T1 for T2–T5 and a treatment T0 without a lead donor for T1.

[4] We do not compare predicted choices with each other.

[5] Notice that an alternative would be to take the entire sample as a denominator (for example, people who always give zero are always consistent). Our more conservative approach adjusts for cases of low power.

than €20 even absent any match, to avoid confounding the comparisons with a change in the identity of the marginal donor. For these donors the treatment corresponds to a *de facto* increase in income rather than a conditional increase in income as they would have donated some positive amount in any case. When focusing on high valuation donors, the number of violations falls considerably. 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 income treatment T5 but that would not have donated in any other counterfactual treatment.

To summarize, the behavior of 88 individuals is predicted to violate revealed preferences (out of 466),[6] while at least 80% of recipients' behavior is consistent with GARP. Whether this is a large or small number depends on the power of our tests, 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 4.

To provide a sense of which of the pairwise comparisons are most informative, we consider the following alternative hypothesis. We generate predicted choices for each donor by first estimating a specification analogous to (2) but excluding the treatment dummy. Column 4 of Table 4 then shows the number and percentage of violations of GARP that would have occurred under this alternative hypothesis. For eight out of the ten pairwise comparisons the number of actual violations is equal or smaller than the number of violations based on this alternative, in some cases by orders of magnitudes, suggesting these pairwise comparisons are powerful tests of GARP. More details of this test are provided in the Technical Appendix.

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. In short, in a real world static environment where participants make simple decisions they are familiar with, the predictions of microeconomic theory work well

[6] Note that some conditions overlap.

in explaining the observed choices of individuals.

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

[4] 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.

[5] 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.

[6] BLUNDELL.R, M.BROWNING, AND I.CRAWFORD (2007) "Improving Revealed Preference Bounds on Demand Responses", *International Economic Review* 48: 1227-44.

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

[8] CHOI.S, R.FISMAN, D.GALE, AND S.KARIV (2007) "Consistency and Heterogeneity of Individual Behavior under Uncertainty", *American Economic Review* 97: 1921-38.

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

[10] DE OLIVEIRA.A.C.M, R.T.A.CROSON AND C.ECKEL (2011) "The Giving Type: Identifying Donors", *Journal of Public Economics* 95: 428–435.

[11] DELLA VIGNA.S (2009) "Psychology and Economics: Evidence from the Field", *Journal of Economic Literature* 47: 315-72.

[12] ECKEL.C AND P.GROSSMAN (2008) Subsidizing Charitable Giving: A Field Test Comparing Matching and Rebate Subsidies", *Experimental Economics* 11: 234–252.

[13] FREY.B.S AND S.MEIER (2004) "Social Comparisons and Pro-social Behavior: Testing 'Conditional Cooperation' in a Field Experiment", *American Economic Review* 94: 1717-22.

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

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

[16] HUCK.S AND I.RASUL (2011) “Matched Fundraising: Evidence from a Natural Field Experiment”, *Journal of Public Economics* 95: 351–362.

[17] HUCK.S, I.RASUL AND A.SHEPHARD (2015): “Comparing Charitable Fundraising Schemes: Evidence from a Natural Field Experiment and a Structural Model”, *American Economic Journal: Economic Policy* 7: 326–369.

[18] KARLAN.D AND J.A.LIST (2007) “Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment”, *American Economic Review* 97: 1774-93.

[19] 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.

[20] LIST.J.A AND D.MILLIMET (2008) “The Market: Catalyst for Rationality and Filter of Irrationality”, *The B.E. Journal of Economic Analysis & Policy (Frontiers)* 8: Article 47.

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

[22] 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.

[23] SIPPEL.R (1997) “An Experiment on the Pure Theory of Consumer’s Behaviour”, *Economic Journal* 107: 1431-44.

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

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

[26] VESTERLUND.L (2003) “The Informational Value of Sequential Fundraising”, *Journal of Public Economics* 87: 627-57.

Table 1: 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 2: 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	Total Amount Raised	Number of Donors	Response Rate	Average Donation Received	Median Donation Received	Average Donation Given	Median Donation Given
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	Lead donor (Control)		17416	17416	132	.035	132	100	132	100
						(.003)	(14.3)		(14.3)	
2	Lead donor + 1:.5 matching					.042	151	75	101	50
			15705	23558	156	(.003)	(18.9)		(12.6)	
		T1				[.134]	[.421]	[.131]	[.102]	[.000]
3	Lead donor + 1:1 matching					.042	185	100	92.3	50
			14310	28620	155	(.003)	(20.7)		(10.4)	
		T1				[.133]	[.037]	[.999]	[.025]	[.000]
		T2				[.994]	[.231]	[.217]	[.609]	[1.000]
4	Lead donor + 1:1 matching for donations greater than €50					.043	194	120	97.9	60
			15671	31107	160	(.003)	(19.3)		(9.59)	
		T1				[.084]	[.010]	[.102]	[.049]	[.000]
		T2				[.820]	[.109]	[.001]	[.863]	[.149]
		T3				[.826]	[.730]	[.260]	[.681]	[.260]
5	Lead donor + €20 match for any donation					.047	89.2	70	69.2	50
			12248	15788	177	(.003)	(5.51)		(5.51)	
		T1				[.008]	[.006]	[.065]	[.000]	[.002]
		T2				[.240]	[.002]	[.751]	[.023]	[1.000]
		T3				[.244]	[.000]	[.008]	[.049]	[1.000]
		T4				[.343]	[.000]	[.000]	[.010]	[.084]

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 then received by the opera house in each treatment is reported in the donation received column. All monetary amounts are measured in Euros.

Table 3: Pairwise Tests of Revealed Preference

Treatments Being Compared		**Type of Comparison**	**Response Rate [One Sided t-test]** (1)	**Proportions Above/Below Some Critical Value [One Sided t-test]** (2)	**Distribution of Donations Given [Mann Whitney test]** (3)
T1: Lead donor (control)	**T2: Lead donor + 1:.5 match**	Budget set expands	Weakly higher in T2 [.933]		
T1: Lead donor (control)	**T3: Lead donor + 1:1 match**	Budget set expands	Weakly higher in T3 [.934]		
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 [.958]		
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 [1.000]	
T2: Lead donor + 1:.5 match	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Weakly higher in T5 [.880]	Proportion of donations < 40 weakly higher in T5; Proportion of donations > 40 weakly higher in T2 [.986]	
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 [.413]		$\boxed{\text{Identical for donations > 50 (if no focal point effects) [.000]}}$
T3: Lead donor + 1:1 match	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Weakly higher in T5 [.878]	Proportion of donations < 20 weakly higher in T3; Proportion of donations > 20 weakly higher in T5 [.988]	
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 [.828]	Proportion of donations < 50 weakly higher in T5; Proportion of donations > 50 weakly higher in T4 [1.000]	

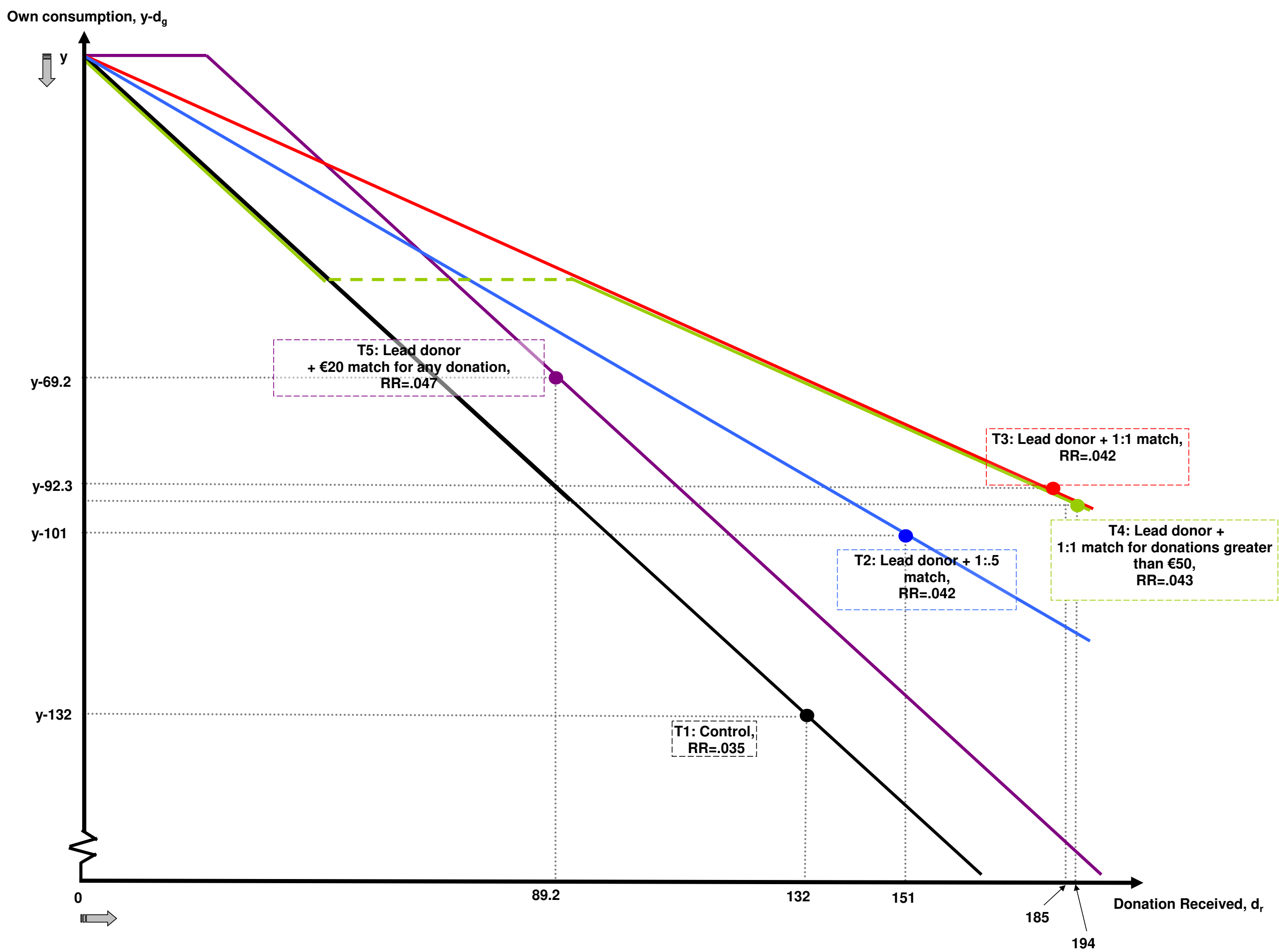
Notes: Hypotheses being tested in columns (1)-(3). They describe behavior that is, on average, consistent with revealed preferences. P-value on relevant test in brackets below. Tests that are outlined are those that require the assumption of convexity on consumer preferences. The tests of proportions are based on all mail out recipients.

Table 4: Individual Violations of Revealed Preference

Matching Treatments Being Compared		Type of Comparison	GARP Violation		Number (Percentage) of Violations	Donation Given Among Violators [95% confidence interval]	Number (Percentage) of Violations, Predicted High Donors	Alternative Hypothesis: Number (Percentage) of Violations
					(1)	(2)	(3)	(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 [N=70]	and predicted to give less than 50 in T4	1 1.4%	49.5		1 1.4%
			Give less than 50 in T4 [N=11]	and predicted to give more than 50 in T1	3 27.3%	52.3 [44.8, 59.7]		8 72.7%
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 [N=62]	and predicted to give less than 50 in T4	2 3.2%	48.2 [38.4, 58.0]		2 3.2%
			Give more than 50 in T4 [N=128]	and predicted to give less than 50 in T2	14 10.9%	44.8 [42.0, 47.6]		35 27.3%
T2: Lead donor + 1:.5 match	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Give less than 40 in T2 [N=48]	and predicted to give more than 40 in T5	46 95.8%	68.0 [63.0, 73.0]	23 47.92%	37 77.1%
			Give more than 40 in T5 [N=103]	and predicted to give less than 40 in T2	0 0.0%	-	0 0.0%	7 6.8%
T3: Lead donor + 1:1 match	**T5: Lead donor + €20 match for any donation**	Budget sets cross	Give less than 20 in T3 [N=15]	and predicted to give more than 20 in T5	15 100.0%	62.3 [53.9, 70.7]	3.00 20.00%	15 100.0%
			Give more than 20 in T5 [N=132]	and predicted to give less than 20 in T3	0 0.0%	-	0.00 0.0%	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 [N=11]	and predicted to give more than 50 in T5	10 90.9%	64.0 [57.8, 70.1]	3.00 27.3%	3 27.3%
			Give more than 50 in T5 [N=55]	and predicted to give less than 50 in T4	0 0.0%	-	0.00 0.0%	0 0.0%

Notes: The number of violations are based on recipients that responded with some positive donation in their assigned treatment.The percentage of violations is the number of violations divided by the number of individuals that fulfills the first part of the condition (N given in square brackets). 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.

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



Notes: Figure 1 graphs the budget sets induced by the five treatments in (y - d_g; d_r)-space. The average in each treatment is marked by a dot on a budget line, and the donation received is marked at the horizontal axis while the donation given is marked at th evertical axis. RR is the response rate in each treatment.

Technical Appendix to: TESTING CONSUMER THEORY: EVIDENCE FROM A NATURAL FIELD EXPERIMENT

MAJA ADENA (WZB) STEFFEN HUCK (WZB & UCL)
IMRAN RASUL (UCL)

SEPTEMBER 2017

Abstract

This appendix contains some additional details of the analysis presented in section 4.2 in the main paper and an additional robustness test. It also includes the mail out letter and its translation.

1 The Prediction Model

It is important to be clear that by using a between-subject design and therefore having to predict recipient behavior to treatments they were not assigned to, the test of GARP partly relies on the accuracy of the prediction model. Table T1 presents the estimates from model (2) that form base for our predictions. Note that the regressions are run pairwise, relative to the control treatment. Out of six regressions only one of the treatment coefficients (on T4) is not significant but it is, at the same time, very small compared to the others.

The potential inaccuracy of the prediction model raises a number of issues. First, it would always be preferable to control for a rich a set of covariates as possible. In this setting, we are restricted to those observables available from the opera house's database that relate to proxies of individual affinity to the opera house and income. Future field experiments should engage in more primary data collection to help address this issue.

Second, the prediction has some associated standard error. Hence although the point estimate of the prediction may suggest a lack of violation of GARP, or *vice versa*, this might not be a particularly informative statement. To see the precision of the prediction, Column 2 also gives the 95% confidence interval for predicted donations among violators of GARP. For example, for the third test based on a comparison of actual behavior in T2 and predicted behavior in T4, the point estimate on the prediction of violators is €48.2 and the 95% confidence interval is from €38.4 to €58. As the confidence interval runs above €50, it might be that all violators actually adhere to GARP. Of course the prediction error also implies that some individuals assigned to adhere to GARP according to their point estimate, might not actually do so at conventional significance levels. In this regard, it is probably helpful to focus on the behavior of a relatively homogeneous group: a specifically targeted group of opera goers, and among this group, those that actually choose to make some positive donation.

Third, the prediction model might be misspecified. As a robustness check, we note that allowing for heterogenous treatment 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, yields similar conclusions. Clearly, more sophisticated approaches could be applied to make predictions that are more robust to functional form misspecifications such as the non-parametric methods developed in Blundell *et al* [2003, 2007].

2 Power

The second set of issues to be discussed with this methodology is in regard to the power of the test. Whether 80% of non-violations of GARP is considered a large or small number depends on the power of our tests, which in turn requires a specific alternative hypothesis to be specified [Varian 1982, Bronars 1995, Andreoni and Harbaugh 2008]. 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. In addition we note that the observables controlled for do have some predictive power in explaining the variation in donations. More precisely we find that in general, recipients that have placed more ticket orders in the 12 months prior to the mail out, and have paid a higher average price per ticket over the same period, donate significantly more regardless of whichever treatment they are assigned to. Hence the observables we condition explain some of the variation in donations, increasing the power of our test to detect violations of revealed preference theory, all else equal.

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 4, but this is a shortcoming of our design that should be borne in mind for the results. This does not however detract from the methodological contribution of our analysis that field experiments can be crafted to test revealed preference theory.[1]

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 (2) but not controlling for any treatment dummy, including the omitted control treatment T1. The results are provided in Table T2. Hence under this alternative we assume donations are driven purely by the observables listed in Table 1 rather than treatment assignment. Column 4 of Table 4 then shows the number and percentage of violations of GARP that would have occurred under this particular alternative hypothesis.

For eight out of the ten pairwise comparisons—except row 5 and 9—the number of violations based on this alternative are always at least as large as the actual number of violations. Note that if the number (percentage) of violations based on the alternative is small, there is not much room for improvement—like in row 10. In some cases, the number of actual 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 14 while 35 violations are predicted under the alternative hypothesis. Similarly, comparing observed donations in T5 and predicted donations in T2, the actual number of violations is 0 while 7 violations are predicted under the alternative hypothesis.

In contrast, a few of the other comparisons distinctly lack power against this specific alternative, which is as expected given the shortcomings arising from the precise location of intersection between budget sets described above. 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 by GARP 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

[1] A series of indices of power of GARP tests are presented in Andreoni and Harbaugh [2008].

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.

Although both methodological issues—the accuracy of the prediction model and power of the GARP test—apply to all empirical approaches, solutions to both might more readily available to field experimenters. In future work using this approach, experimenters need to engage in data collection and design interventions that improve the accuracy and robustness of the prediction model, and allow for more powerful tests of GARP by engineering budget line intersections at bundles closer to the expected behavior of more individuals.

3 Focal Point Effects?

The results in Table 4 highlight that the pairwise comparisons that yield the highest percentage of violations all involve the non-convex treatment T4. As discussed earlier, this might be because individuals that would have given less than €50 in T1 choose to donate slightly more than revealed preference theory predicts in T4 and so do locate just above the interior corner solution of €50 in T4. Moreover, the wording in T4 in the mail out letter might lead to d_g = €50 becoming focal for 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 we use quantile regression methods to characterize the effect of being assigned to treatment T4 relative to T3 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 (2). We therefore estimate the following quantile regression specification at each quantile $\theta \in [0, 1]$,

$$Quant_\theta(\log(d_{gi})|.) = \beta_\theta T4_i + \gamma_\theta X_i \text{ for } d_{gi} > 0. \quad (1)$$

The parameter of interest, β_θ, measures the difference at the θth conditional quantile of log donations received between the treatment group T4 relative to the omitted group T3. Figure 2B graphs estimates of β_θ from (1) and the associated 95% confidence interval at each quantile. We also show the quantile that corresponds to donations given of €50 (€60) across T3 and T4.

Figure T2 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 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.[2]

References

[1] ANDREONI.J AND W.T.HARBAUGH (2008) Power Indices for Revealed Preference Tests, mimeo, UCSD.

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

[3] BLUNDELL.R, M.BROWNING, AND I.CRAWFORD (2007) "Improving Revealed Preference Bounds on Demand Responses", *International Economic Review* 48: 1227-44.

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

[5] 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.

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

[2]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].

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.

Table T1: Estimates of the Model (2)

Dependent variable: Donation given[(a)]

Baseline	T0	T1	T1	T1	T1	T1
Treatment	T1	T2	T3	T4	T5	T5 high donors
Treatment dummy	0.366***	-0.308**	-0.323***	-0.054	-0.505***	-0.324***
	(0.122)	(0.134)	(0.120)	(0.107)	(0.117)	(0.104)
Female dummy	0.066	-0.203	0.111	-0.046	-0.032	0.094
	(0.124)	(0.131)	(0.120)	(0.107)	(0.113)	(0.103)
Number of ticket orders in last 12	0.028**	0.037***	0.028**	0.014	0.028**	0.030**
months	(0.013)	(0.014)	(0.011)	(0.011)	(0.013)	(0.013)
Average value of tickets	0.007***	0.009***	0.008***	0.007***	0.008***	0.007***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
Munich postcode (yes=1)	0.102	0.192	0.161	0.143	0.206*	0.200*
	(0.136)	(0.132)	(0.122)	(0.106)	(0.117)	(0.107)
Dummy=1 if year of last ticket	0.117	0.219	0.027	0.128	0.231*	0.044
purchase=2006, =0 if earlier	(0.146)	(0.153)	(0.146)	(0.129)	(0.130)	(0.116)
Constant	2.987***	3.143***	3.288***	3.518***	3.219***	3.593***
	(0.220)	(0.253)	(0.209)	(0.222)	(0.228)	(0.201)
Observations	274	288	287	292	309	251
Adjusted R^2	0.103	0.114	0.158	0.077	0.149	0.131

Notes: Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, (a) the dependent variable is transformed as follows: f(X)=ln(X-k), with k chosen such that the distribution is symmetric around the mean.

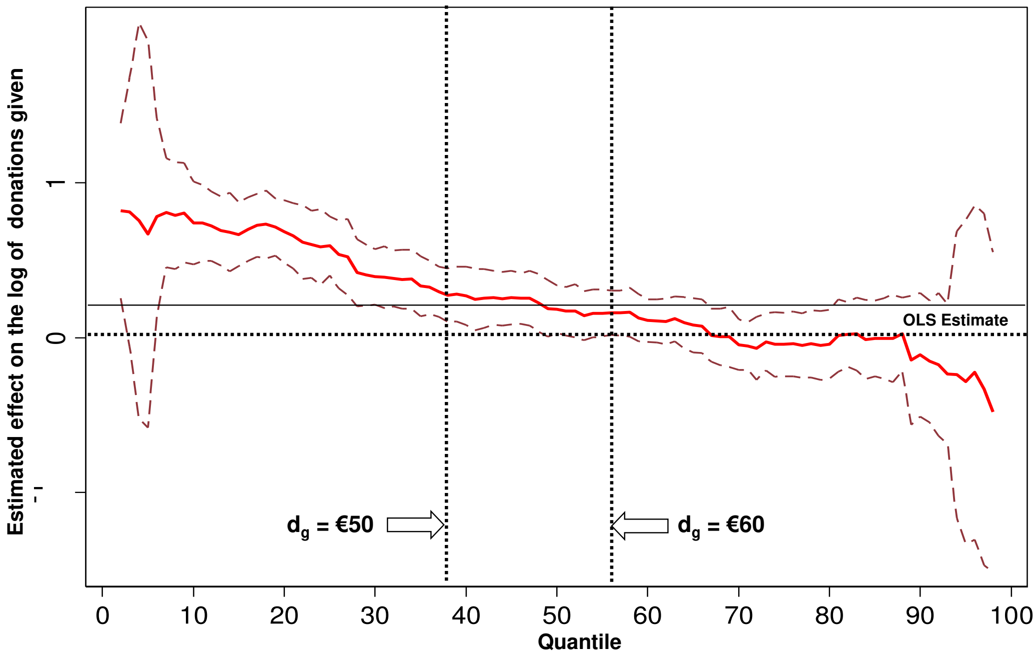
Table T2: Estimates of the Model (2) without treatment dummies

Dependent variable: Donation given[(a)]

Baseline	T0	T1	T1	T1	T1
Treatment	T1	T2	T3	T4	T5
Female dummy	0.115	-0.182	0.143	-0.047	-0.004
	(0.126)	(0.131)	(0.120)	(0.107)	(0.116)
Number of ticket orders in last 12	0.032**	0.039***	0.031***	0.015	0.032**
months	(0.013)	(0.014)	(0.011)	(0.011)	(0.012)
Average value of tickets	0.007***	0.009***	0.009***	0.007***	0.009***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Munich postcode (yes=1)	0.094	0.197	0.149	0.137	0.159
	(0.139)	(0.133)	(0.123)	(0.105)	(0.120)
Dummy=1 if year of last ticket	0.134	0.249	0.020	0.129	0.275**
purchase=2006, =0 if earlier	(0.147)	(0.151)	(0.145)	(0.129)	(0.132)
Constant	3.102***	2.921***	3.094***	3.487***	2.858***
	(0.222)	(0.218)	(0.199)	(0.208)	(0.209)
Observations	274	288	287	292	309
Adjusted R2	0.075	0.099	0.139	0.080	0.096

Notes: Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, (a) the dependent variable is transformed as follows: f(X)=ln(X-k), with k chosen such that the distribution is symmetric around the mean.

Figure T1: Non Convex Budget Set T4 Relative to 100% Match Rate Treatment T3



Notes: The figure 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.