

NBER WORKING PAPER SERIES

MATCHING AND CHALLENGE GIFTS TO CHARITY: EVIDENCE FROM LABORATORY AND NATURAL FIELD EXPERIMENTS

Daniel Rondeau
John A. List

Working Paper 13728
<http://www.nber.org/papers/w13728>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2008

We are grateful to the British Columbia Chapter of the Sierra Club of Canada for allowing us to design their fundraising campaign. Part of the writing of this paper was done while Daniel Rondeau was an OECD fellow and Visiting Scholar at the Laboratoire Méditerranéen d'Économie Théorique et Appliquée (LAMETA), INRA-Montpellier, France. Their financial support and hospitality is gratefully acknowledged. We also thank seminar participants at Wharton, Columbia, and the 2004 EAERE and 2005 ASSA meetings. Richard Carson, Rachel Croson, Glenn Harrison, Liesl Koch, David Reiley, Lise Vesterlund, Stefano della Vigna, Timothy Cason and three anonymous referees provided remarks that improved the manuscript. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2008 by Daniel Rondeau and John A. List. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Matching and Challenge Gifts to Charity: Evidence from Laboratory and Natural Field Experiments
Daniel Rondeau and John A. List
NBER Working Paper No. 13728
January 2008
JEL No. C9,C93,H4

ABSTRACT

This study designs a natural field experiment linked to a controlled laboratory experiment to examine the effectiveness of matching gifts and challenge gifts, two popular strategies used to secure a portion of the \$200 billion annually given to charities. We find evidence that challenge gifts positively influence contributions in the field, but matching gifts do not. Methodologically, we find important similarities and dissimilarities between behavior in the lab and the field. Overall, our results have clear implications for fundraisers and provide avenues for future empirical and theoretical work on charitable giving.

Daniel Rondeau
Department of Economics
University of Victoria
Victoria, BC, Canada V8W 2Y2
rondeau@uvic.ca

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@arec.umd.edu

1. Introduction

Charitable fundraising remains an important matter for the international community and more narrowly in the U.S., where the American Association of Fundraising Counsel estimates that total contributions to American philanthropic organizations in the year 2000 exceeded 2 percent of GDP. Dove (2000) reports that there is at least one capital campaign under way in virtually every major population center in North America that has an objective of raising between \$25 million and \$100 million. Smaller capital campaigns are even more numerous.

Fundraisers have developed a variety of strategies aimed at successfully reaching specific funding targets, but to date, little is known empirically about the demand side of charitable fundraising. For example, without more than a handful of anecdotes, professional fundraising consultants strongly assert that “leadership gifts” are an important step towards a broader public campaign.¹ Such leadership contributions typically take one of two forms: challenge gifts and matching gifts.

A challenge gift is an *unconditional* commitment by a donor, or set of donors, to provide a given sum of money to the cause. Professional fundraising consultants advise seeking at least 10% and up to 80% of the overall objective of a capital campaign through leadership gifts before embarking on the public phase of fundraising (Lawson, 2001). Alternatively, a matching gift is a *conditional*

¹ Leadership gifts should be distinguished from rebate mechanisms such as tax deductions, whereby charitable contributions can be used to reduce one’s tax burden. There is an extensive literature estimating price elasticities of such government-sponsored tools (see, e.g., Randolph, 1995; Auten et al., 2002), and it is fair to say that the empirical estimates vary widely.

commitment by the leadership donor(s) to match the contributions of others at a given rate, up to the maximum amount the leadership donor is prepared to give. While the rate of matching is typically the result of an agreement between the fundraiser and the leadership donor, received wisdom suggests that presenting less than a 1:2 matching ratio (one leadership dollar for every two dollars contributed by other donors) has noticeably less power to leverage other contributions than more generous matching rates (e.g., 1:1 or 2:1) (Dove, 2000).

Interestingly, these “rules of thumb” are largely anecdotal. Only recently have economists begun to formally examine such claims. For example, List and Lucking-Reiley (2002) find that challenge gifts announced by the fundraiser influence charitable contributions. Landry et al. (2006) report data from a door-to-door fundraising drive that suggests challenge gifts increase conditional contributions. Meier (2007) finds that a matching contribution has an immediate positive effect on charitable contributions, but notes that the net long run effect is to reduce the overall level of contribution of the group when individuals are given repeated opportunities to give. Karlan and List (2007) also obtain mixed results on the effect of matching gifts: matching is shown to have a significant effect on the contributions of some but not all groups of potential donors. The relative efficacy of challenge and matching gifts, however, has not to our knowledge been studied systematically, neither in the field nor in the laboratory.²

² In a much different environment and context, Eckel and Grossman (2003) compare matching to an equivalent rebate of one’s contributions in a laboratory dictator game and find that matching contributions lead to significantly larger contributions than the rebate mechanism. The interested

Our first objective is to perform such a comparison in a natural field experiment (see the classification scheme in Harrison and List, 2004) using four parallel fundraising appeals conducted as part of the British Columbia Chapter of the Sierra Club of Canada's normal fundraising efforts. Using direct mail solicitation, 3,000 Sierra Club supporters were randomly divided into four treatments and asked to support the expansion of a K-12 environmental education program. We designed the experiment in an effort to provide a clean test of the absolute and relative efficacy of challenge and matching gifts.

We report some potentially useful insights. We observe that a challenge gift attracted 23% more donors and increased total dollar contributions 18% when compared to the identical campaign in which no announcement of leadership gift was made. While the number of donors was also higher in the matching treatment, average and total donations under matching were actually *lower* than in the challenge and control treatments. We also find that increasing the monetary target of the fundraising campaign has a significant effect on the size of the average gift.

Our second objective is to compare these insights with those gained from a controlled laboratory experiment. The combination of lab and field data provides a deeper understanding than either approach could provide in isolation. Such an exercise also permits an exploration of an important issue facing experimental economists: whether behavior inside the laboratory is a good indicator of behavior outside the laboratory.

reader should also see Frey and Meier (2004), Shang and Croson (2005), Falk (2007), or the survey on field experiments on charitable giving in List (2006).

We report empirical results from five laboratory treatments, revealing a mix of similarities and differences with data from the field experiment. One difference is that contrary to our observations from the field, the rate of participation in the laboratory (giving to the public good) is nearly 100%—i.e., a strong majority of subjects give something to the public good. A second result is the close correspondence between the relative rankings of monies raised in the lab and the field: in terms of relative mechanism performance, the lab and field data share strong similarities. Third, while the field results show that donors respond strongly to the change in the cost of the program, the laboratory treatments offer mixed evidence on the extent of this effect. Combining insights from the lab and the field provides evidence consistent with the notion that solicitees in the field viewed the stated monetary goal as a signal of the quality of the public good to be provided.

2. Controlled Fundraising Campaigns in the Field

2.1 Design

In June 2004, the British Columbia chapter of the Sierra Club of Canada mounted four parallel fundraising campaigns in an effort to expand their K-12 environmental education programs. This chapter of the Sierra Club typically appeals to its members and supporters to send tax-deductible donations three to four times per year. Our field experiment was one of those fundraising drives, and the basic design is inspired by both fundraising practices and previous laboratory threshold public goods experiments (for recent examples, see Rondeau et al. 1999; List and Rondeau, 2003, Rondeau et al., 2005).

A total of 3,000 Sierra Club supporters were randomly divided into four treatments. Supporters received a solicitation letter from the director of the Club, written, as usual, on Sierra Club letterhead and accompanied by a payment return card of a format typically used by the Club in its fundraising. The letter explained the Club's objective of expanding its existing K-12 environmental education program to increase the number of students it reaches by at least 200. We designed our solicitation as a threshold public good with a money-back guarantee. All pledge letters specified that 750 supporters were being solicited and that the program would be expanded only if a specified minimum amount of money was raised. In three of the treatments, the total amount of money required was set at \$5,000. In the fourth treatment letter, the announced threshold was \$2,500.

Importantly, in all cases, it was clearly indicated that if the minimum amount required was not received, individual contributions would be returned to donors. The inclusion of a money-back guarantee in a threshold public good campaign is not an everyday occurrence, but motivated by other uses in the field (e.g., Manitoba's New Democratic Party, the Association of Oregon Faculties, and Office coffee clubs—see List and Lucking-Reiley) and the findings of Rondeau et al. (1999, 2005), which reveal that under a range of conditions, the combination of a threshold and money-back-guarantee significantly increases contributions in laboratory experiments.

More importantly, this design choice provides a level playing field for the relative test of efficacy across the challenge and matching gift treatments under a theory of signaling (Vesterlund, 2003; Andreoni, 2006). This intuition holds because if leadership gifts actually signal the type and value of the public good, they should increase contributions relative to the case with no announced leadership gifts.³ With the inclusion of a money-back guarantee, both challenge

³ One simple mechanism underlying this effect is that when a large donor believes in the project by giving a gift this informational value affects an agent's beliefs regarding the value of the public good. Andreoni (1998) discusses a different effect of challenge monies: his model of charitable giving for a threshold public good has multiple equilibria, and in the absence of seed money there exists a Nash equilibrium with zero charitable giving. The zero-contribution equilibrium can be eliminated, however, by initial commitments of seed money, which lower the remaining amount needed to be raised in the public fundraising campaign. Thus, in his model seed money is used as an elimination device rather than as a credibility device.

gifts and matching gifts yield equivalent outcomes whether or not the threshold goal is reached.⁴

Solicitation letters also explained that any amount of money collected in excess of the threshold would serve to further increase the number of students reached by the program (at the rate of one additional classroom for each \$775 received). The four treatments were:

- a “High Control” (HC) group, in which the full amount of \$5,000 is solicited from the members receiving the letter;
- a “Challenge” (CH) treatment, in which \$5,000 has to be raised, but where a leading donor has already committed \$2,500 in a challenge gift to the cause;
- a “Matching” (MA) treatment, where \$5,000 has to be raised, but where a leading donor has committed to match the first \$2,500 in contributions at a rate of 1:1;
- a “Low Control” (LC) treatment, in which the existence of a leadership gift is not announced but where \$2,500 is requested from donors.

We provide verbatim copies of the letters in the supplementary material found on the journal’s website. Key portions of the payment mechanism and treatment variations are reproduced here:

This year one of the main goals of the Sierra Club is to expand the scope of our environmental education program. This is why you are one of 750 Sierra Club supporters receiving this letter.

⁴ For example, if the threshold (goal) is not reached, all donations from the solicitees are refunded in both cases. Alternatively, if the goal is reached, in each case the full amount of the leadership gift is used to help fund the public good. Thus, the money-back guarantee serves to remove the conditionality inherent in the matching gift solicitation if the goal is not reached. The money-back guarantee therefore serves to equalize the price of giving over all ranges of giving for this particular public good. An astute reader might wonder in practice how a money-back guarantee influences the leadership gift. Following standard practices, a challenge gift remains with the charity but are used for other undefined purpose. Alternatively, the matching monies are not received.

If we raise \$5,000, we could prepare and deliver our education program to an additional 200 children across the province. *So we are appealing to donors like you to help us raise the \$5,000 needed to expand the program.*

But wait, we can do even more! If we receive more than \$5,000 from this group of 750 supporters, we will be able to reach an even greater number of students. **With each extra \$775 we can deliver the program to one additional classroom.**

On the other hand, if we fail to raise \$5,000 from our supporters we will not be able to bring the gift of environmental education to more children this year, and we will refund your donation to you.

In the LC treatment, the letter was in all respects identical except that all instances of “\$5,000” were replaced by \$2,500.⁵ In the matching treatment, the second paragraph above became:

If we raise \$5,000, we could prepare and deliver our education program to an additional 200 children across the province. The **great news** is that we have already received a commitment of \$2,500 in matching donations. **This means that for each of the first \$2,500 we receive from supporters like you, we will also receive another dollar and double the impact of your donation.** *So we are appealing to donors like you to help us raise the \$2,500 needed to expand the program.*

In addition, the residual amount of money required mentioned in the last two paragraphs (\$5,000) was changed to \$2,500. Similar modifications were made for the challenge treatment:

If we raise \$5,000, we could prepare and deliver our education program to an additional 200 children across the province. The **great news** is that we have already received a commitment of \$2,500 in start-up money. *So we are appealing to donors like you to help us raise the \$2,500 needed to expand the program.*

⁵ The Sierra Club committed the additional \$2500 necessary to reach 200 children if they raised \$2,500 in this treatment.

Together, these four treatments represent experimental cells that, to our knowledge, have not been explored heretofore and allow for the testing of hypotheses: i) the size of a threshold affects giving; ii) announcing a lead gift increases contributions; and iii) the form of the lead gift will affect contributions.

2.2 Results

A quick overview of the results can be gleaned from the summary data shown in Table 1.⁶ One clear pattern is that the challenge (CH) treatment raises considerably more money than the other three treatments: the \$1,620 raised is 18%, 31%, and 71% higher than the total amount raised in the HC, MA, and LC treatments, respectively. The donation of those who answered the different appeals (overall mean, conditional on giving) varies from \$45.83 in the HC treatment to a low of \$27.79 in the LC group. The average donation in the CH treatment, at \$43.78, is comparable to that of the HC treatment, whereas the average donation in the MA treatment (\$34.31) is closer to the amount observed in the LC treatment. Considering mean conditional contributions, Tables 2 and 3 provide the results of pair-wise tests of equality of the mean donations and of homogeneity of the distributions.

--- INSERT TABLES 1, 2 and 3 HERE ---

Examination of the data reveals that (1) there is a strong threshold effect—decreasing the announced cost of the good significantly decreases contributions; (2) consistent with anecdotal evidence and past field results, we find that the use

⁶ Our results concerning the challenge treatment would be stronger, but we excluded one contribution of \$400 from the challenge treatment because the donor later retracted her check citing that she had made a mistake when contributing.

of challenge funds has a positive influence on contributions—the challenge treatment yielded a participation rate 23% (9%) higher and overall contributions 18% (71%) higher than the High (Low) Control treatment; and (3) matching donations provides no advantage to the fundraiser, yet monies raised are not significantly lower than the challenge treatment. We consider each result in turn.

2.2.1 Strong Threshold Effect in the Field

The HC treatment raised \$1,375, whereas the LC treatment raised \$945. Neither treatment mentions a leadership gift. The only difference between the two appeals is the minimum amount of donations required to expand the education program (\$5,000 in the HC treatment versus \$2,500 in the LC treatment).

Since the LC treatment induced a slightly higher response rate, the observed difference in the total amount received is attributable to larger individual donations in the HC treatment. The difference between the mean contribution of \$45.83 in the HC treatment and \$27.79 in the LC treatment represents an increase of roughly 65%, and is statistically significant at the $p < 0.02$ level using a two-tailed t-test (A generalized rank-sum test of distribution homogeneity rejects the null hypothesis that the two distributions are identical at the $p < 0.08$ level).⁷

⁷ This generalized rank-sum test, which is described in Hoel (1971), is less restrictive than other non-parametric tests (e.g., a Mann-Whitney test) in that it requires no assumption about the shape of the underlying distribution being tested.

These results suggest that the size of contributions is positively correlated with the threshold level.⁸ Such a finding is consistent with the notion that individual donors place a greater weight on the expressed need of the Sierra Club than on the “price” of the education program (or the benefit-cost ratio of the program, which was twice as large in the LC treatment compared to the HC treatment).

Despite the fact that Sierra Club members did not have information about the value of the public good to others, the observed threshold effect is also consistent with the Nash prediction for contributions in a provision point game with complete information (Bagnoli and Lipman, 1989). In the efficient equilibria of this game, the sum of contributions equals the cost of the good (and each individual contributes no more than his value). Incentives to free-ride remain, however. Thus, the theoretical prediction implies an increase in the average contribution with the increase in the threshold provided the threshold remains lower than total benefits. Though we observe the static result, group contributions did not reach the threshold.⁹

2.2.2 The Challenge Gift Increases Total Contributions

⁸ As an astute reviewer noted, this cannot be a general result since an enormous goal would likely lead to few contributions. We concur with this intuition.

⁹ The \$5,000 threshold level had been conservatively chosen by the Sierra Club on the basis of their historical records on donor rate of response (13%-14%) and average donation (greater than \$50). This fund raising effort produced much lower response rates and contributions. However, a regular fund raising effort conducted simultaneously by the Club with a different subset of its members produced an even smaller rate of response and smaller contributions. Club staff has no explanation for the dismal results—none of our treatments reached the threshold—but neither we nor they attribute it to our experimental design.

The announcement of a challenge gift yielded \$1,620 compared to the \$1,375 raised from the HC group. This difference in overall contributions is not due to an increase in average contributions, as the average gift conditional on giving was \$43.78 and \$45.83 in the CH and HC treatments, respectively. Rather, it is driven by the 23% difference in the participation rates across the two treatments and the fact that the number of “large” gifts (gifts of \$50 or more) is 27% greater in the challenge treatment. Both of these results are consistent with the empirical results of List and Lucking-Reiley (2002).

Further insights on the usefulness of challenge gifts can be obtained from comparing contributions when a challenge gift is announced (CH treatment) to outcomes of a campaign where an amount equivalent to the challenge gift has been secured by the fundraiser but not announced. The LC treatment permits this comparison. In this case, the effective amount of the threshold is simply reduced by the amount of the implicit challenge gift.

When comparing contributions in the LC and CH treatments the announced cost of the good varies between treatments, but the amount of money required of the solicited individuals remains constant at \$2,500. This comparison therefore provides another measure of the value of the signal sent by the challenge gift. We find that the average contribution increases from \$27.79 in the LC treatment to \$43.78 when a challenge gift is announced, a difference that is statistically significant at the $p < .02$ level using a two-tailed t-test ($p < 0.04$ using the generalized rank-sum test). The total amount raised also increases 71 percent,

from \$945 to \$1,620, due not only to higher average gifts, but also to a slightly greater number of contributors (though the difference in the number of contributors - 37 versus 34 is not statistically significant via a test of proportions on the response rate). We believe that this experimental comparison is novel to the literature, and represents an important piece of the puzzle to interpret previous data (e.g., List and Lucking-Reiley. 2002).

2.2.3 The Matching Gift Does Not Increase Contributions

The data suggest that the gains from implementing a matching gift campaign are not as strong as the gains from a challenge gift campaign. While more funds were raised in the MA than in the LC treatment (\$1,235 versus \$945), the matching contributions are considerably lower than the \$1,620 raised in the CH treatment. Furthermore, the average contribution from matching donors is not statistically different from any of the other three treatment means at conventional levels. These results provide little support for the received fundraiser view that announcing that individual donations will be matched on a 1:1 ratio by a leader is an effective way of leveraging higher donations in the public phase of the campaign. While the total amount contributed in the CH treatment exceeds that contributed in the MA treatment, none of our statistical tests can reject the null that the differences are due to noise.

This result differs with those of Karlan and List (2007) and of Meier (2007), who find that matching significantly increases contributions. Our respective results are not directly comparable, however, since their campaigns did

not include a threshold, nor a money back-guarantee. The voluntary contribution and threshold mechanisms produce radically different incentives that might explain the difference, but it is worth noting that both studies show that the impact of matching is not straightforward. Karlan and List show that it increases donations only in some geographical areas, and Meier finds that those who give under a matching condition later reduced their contributions to the same cause, resulting in a net reduction of total contributions. Much remains to be learned. In this spirit, we now turn to a set of parallel laboratory experiments.

3. Laboratory Experiment

While the above insights are useful for both positive and practical purposes it is interesting and relevant to consider whether a laboratory experiment with student subjects can yield similar quantitative and qualitative insights.

There are reasons why the insights gained in the lab and in the field might be at odds. A recent framework by Levitt and List (2007) highlights that to the extent that lab and naturally-occurring environments differ on any of several dimensions—e.g., the nature and extent of scrutiny, the particular context in which a decision is embedded, and the manner in which participants are selected to participate—the results obtained across domains might differ. This is not to suggest that either domain is inferior in some manner, rather we aim to explore whether there are differences; and, if so, what causes such behavioral differences.

3.1 Design

In designing the laboratory experiment a first difficult question is whether to parallel the field treatments exactly. An alternative approach is to hold true to the experimental literature and set-up an environment to replicate the important features of the theory while maintaining a connection to the field environment of interest. We opted for the latter because we also aim to speak to the generalizability of extant laboratory results, which are typically designed to replicate the important features of the theory. Thus, the design of our laboratory experiment closely resembles previous threshold public goods experiments (Rondeau et al. 1999, 2005; List and Rondeau, 2003).

Subjects in groups of six are given an initial endowment of \$12 and asked to divide this amount between a private account and a group fund. Any division is allowed, provided the amounts in the two accounts are positive and sum to exactly \$12. An investment is made by the group fund only if the sum of contributions by group members reaches a minimum threshold (the provision point). If the group fails to reach the threshold, all contributions are refunded. If the threshold is reached or exceeded, a fixed payoff (of \$4.50 or \$9.00 depending on the treatment) is paid to each member of the group. In addition, each group member receives an additional payoff for contributions in excess of the threshold. This additional payoff is paid at the rate of \$0.20 per dollar of excess contributions.

In all treatments, everything about the experiment was common knowledge with the exception of the payoffs of other players. While everyone had

identical payoffs, subjects were told that they were randomly assigned to their payoff amounts and that others *may not* have the same payoff amount.¹⁰

--- INSERT TABLE 4 HERE ---

Table 4 summarizes the experimental design.¹¹ In the High Control (HC) treatment, the public good was provided if a minimum of \$45 was contributed to the group fund. If the good was funded, subjects received \$9 as a fixed payoff plus 0.2 times the amount of excess contributions. In the Challenge Gift (CH) treatment the only change is that subjects were told that an amount of \$22.50 had already been committed to the group fund by “an outside investor.” The Matching Gift (MA) treatment was identical to the HC and the CH treatments except for the fact that the first \$22.50 placed in the group investment fund would be matched by an equal investment (a matching ratio of 1:1) from “an outside investor.”

The Low Control (LC) and Low Control 2 (LC2) treatments provide useful controls for the leadership treatments and lend further insights into the influence of the threshold in these games. The LC treatment had a threshold of \$22.50, benefits of \$9 per subjects at the threshold, and additional payoffs of

¹⁰ One referee was of the opinion that since all induced values were identical, our statements to subjects were a mild form of deception. In fact, subjects were randomly assigned to a seat with a set of instructions when they entered the room; we did not say that the values were randomly chosen; and while all induced values were equal, the statement that other subjects “may not” have the same amounts does not rule out that they may.

¹¹ All five treatments were conducted with student subjects recruited voluntarily from first year introductory economics classes at the University of Victoria. HC, CH and MA are direct parallels to three of the four field treatments. The LC treatment matches a field treatment but these data were collected at a different time (i.e., with a different cohort of similarly recruited first and second year economics students). While we have no reason to believe that systematic differences exist between subjects at these two points in time, we cannot definitively assert that this has no influence.

\$0.20 per dollar in excess of the threshold. Alternatively, LC2 had a provision point of \$22.50, leading to fixed payoffs of \$4.50, and additional payoffs of \$0.20 per dollar in excess of the threshold. The LC2 treatment therefore corresponds to a situation where half the money is sought in order to deliver half of the benefits.

The experiment was implemented as a single-shot game, with students earning approximately CA\$15 on average for a 45-minute session. At the time of recruitment, students were told that they would have an opportunity to earn money in research experiments completely unrelated to the course they were enrolled in and that no knowledge of economics was required to participate. Upon arrival at a predetermined meeting room, students were asked to read and sign a consent form. Communication between students from this point on was strictly forbidden. Written instructions were distributed, read by subjects and summarized orally. Individual questions were answered when necessary.

In sum, a total of 240 students participated in the experiment in sessions of 12, 18, or 24. All five treatments were conducted with student subjects recruited voluntarily from introductory economics classes at the University of Victoria. Treatments HC, CH, MA, and LC are direct parallels to the four field treatments. Treatment LC2 is a variation for which we do not have a parallel field treatment since its low threshold and low benefits structure made it impractical for the Sierra Club.

3.2 Results

Tables 5 and 6 summarize our laboratory data. Contributions to the public good amount to roughly 50% of individual endowment, leading to positive provision of the public good in every treatment. This result is consistent with previous laboratory studies.

The average contribution level in the HC, \$7.53, is higher than the CH treatment mean of \$6.84, the MA treatment mean of \$5.76, and the LC treatment mean of \$7.33. Statistical tests on these means (and medians) yield four observations:¹² (1) there is a threshold effect—decreasing the amount of money required to fund the public good strongly decreases contributions; (2) subjects have a strong positive reaction to the level of the induced value; (3) neither leadership contribution treatment raises more money than the HC treatment; and (4) the CH treatment raises more funds than the MA and LC treatments. In the following sub-sections, we review each result in more detail.

---- INSERT TABLES 5 and 6 HERE ----

3.2.1 Strong Threshold and Strong Induced Value Effects

Recall that the HC and LC2 treatments are identical except for the threshold level of contributions at which subjects begin to receive a return from the group fund. This design choice ensures that the benefit-cost ratio of any sum of contributions leading to the provision of the public good was constant at 1.2.

The comparison of these two treatments, therefore, allows us to isolate the effect

¹² Jarque-Berra tests do not reject the null that the data for each treatment is normally distributed. Pairwise tests also fail to reject equality of variances. As a result, we only report t-tests on the experimental data.

of simply reducing the level (from \$45 to \$22.50) at which individuals start obtaining benefits from the (otherwise linear) public good.

We find that this reduction in the threshold reduced the mean contribution by more than 30%: from \$7.53 in the HC to \$5.41 in the LC2 treatment, and that this difference is highly significant in a two-tailed t-test ($p < 0.002$). This is consistent with a strong and significant threshold effect and the results of Cadsby and Maynes (1999), who report that contributions in a similar threshold game are positively correlated with the provision point.

The LC data can be compared to the LC2 to gauge the effect of doubling the personal benefits while the threshold remains constant at \$22.50. Doing so reveals that subjects are also responsive to the benefits they stand to receive from the public goods. Doubling the induced value increased contributions from \$5.41 to \$7.33, a difference that is significant at $p < 0.01$.

This result is consistent with the findings of Rondeau et al. (2005) who show that the contributions of laboratory subjects in similar provision point experiments (but where subjects have less information than in our case) respond positively and significantly to increases in induced value. They also show that the elasticity of this response is less than unity, which is also consistent with our results. It is also worth noting that the magnitude of the threshold and induced value effects essentially cancel each other out in our experiment (the equality of the HC and LC means cannot be rejected: $p > 0.74$).

The key difference between the lab and field results is that the field HC and LC contributions are significantly different while in the lab, HC and LC are indistinguishable, both higher than LC2. Thus, in the lab, subjects respond to both a decrease in the threshold provision cost, and to an increase in its benefit-cost ratio in a manner that cancels out the two opposing effects. In the field, we observe a drop in contributions associated with the lower threshold, but no opposite effect. This suggests that field participants put a greater weight on the threshold cost than subjects in the laboratory experiment. This, we surmise, might be explained by field participants taking the stated cost as a signal of the quality of the good to be provided. This might be so because the value of the field public good holds a far greater degree of uncertainty than that of the laboratory public good.

3.2.2 Challenge and Matching Gifts Do Not Increase Contributions in the Laboratory

The CH and Ma treatments each raised smaller contributions per subject than the HC or LC treatments. For the MA treatment, those differences are statistically significant ($p < 0.004$ that MA=HC; and $p < 0.02$ that MA=LC). While the CH treatment raised more money than the MA treatment, the difference lies outside of the normally accepted level of statistical significance ($p > 0.11$). Thus, taken alone, the laboratory experiment provides little support for the view that

leadership gifts increase the amount of funds raised.¹³ It is worth keeping in mind that in our laboratory environment where individual benefits are known with certainty, a leadership gift does not carry any information on the quality of the public good. This is a key difference between the lab and field. Given this difference and the higher observed field contributions in the Challenge gift treatment, one possible explanation (for the difference in difference) is once again that the leadership gift plays a signaling role of the value of field public goods.

At the same time, it is interesting to note that the field and laboratory average contribution orderings of the HC, MA, and CH treatments generated identical rankings. This tends to support of the view that despite the qualitative differences between field and laboratory leadership gifts, the laboratory is potentially a useful testbed for fundraising institutions. Most importantly, this provides one piece of empirical evidence that the qualitative, or directional, insights gained from a generic lab experiment can be a useful indicator of treatment effects in naturally-occurring environments.

4. Concluding Remarks

The popular literature on the demand side of charitable fundraising is long on advice and accepted wisdom but largely unsubstantiated by “hard facts”. Controlled experiments provide a means to explore questions heretofore unanswered within the area of charitable fundraising. In this study, we directly pit challenge gifts and matching gifts in an effort to advance our understanding of the

¹³ An interesting point of future research is to compare behavior in laboratory experiments where values are uncertain, but linked to the level of the leadership gift.

optimal uses of leadership gifts. Our data provide some evidence on the power of challenge gifts in the field: they work to increase the size of individual contributions relative to the comparable treatments in which the same amount is required but no challenge is present. They may also induce greater participation.¹⁴

We view our set of experiments as also making a broader methodological point and playing a dual role in the research discovery process. First, field experiments are capable of providing evidence on the generalizability of laboratory evidence. Importantly, we find that the qualitative insights gained from the laboratory are quite similar to those found in the field. Second, our lab experiment helped to uncover the causes and underlying conditions necessary to produce the data patterns observed in the field. Viewed in this light, our results should have practical import for fundraisers as well as positive and normative implications for economists. Our contribution also highlights the complementarities that lab and field experiments provide, as stronger inference can be made when combining the most attractive features of each approach.

¹⁴ While this latter effect might not be as important when soliciting from a “warm” list of contributors, it is invaluable when soliciting from a “cold” list since fundraisers understand that attracting a first donation from an individual is also identifying a new supporter, a crucial element in ensuring the long-term viability of a charity. Indeed, fundraising strategists typically rank building a “donor development pyramid” as the most important aspect of a successful long-term fundraising effort.

References

- Andreoni, J. (2006). Leadership Giving in Charitable Fundraising. *Journal of Public Economic Theory* 8(1): 1-22.
- Andreoni, J. (1998). Toward a Theory of Charitable Fund-Raising. *Journal of Political Economy*, 106(6): 1186-1213.
- Auten, G., H. Sieg, C.T. Clotfelter. (2002). Charitable Giving, Income and Taxes: An Analysis of Panel Data. *American Economic Review*, 92: 371-82.
- Bagnoli, Mark, and Barton L. Lipman. (1989). Provision of Public Goods: Fully Implementing the Core through Private Contributions,” *Review of Economic Studies*, 56: 583-601.
- Cadsby, C.B. and E. Maynes. (1999). Voluntary provision of threshold public goods with continuous contributions: experimental evidence, *Journal of Public Economics* 71, 53-73.
- Dove K.E. (2000). *Conducting a Successful Capital Campaign*. 2nd edition. San Francisco, Jossey-Bass. 510pp.
- Eckel, Catherine C. and Phillip J. Grossman. (2003). “Rebate versus Matching: Does How We Subsidize Charitable Contributions Matter?,” *Journal of Public Economics*, 87, 681–701.
- Falk, Armin. (2007). “Gift-Exchange in the Field.” *Econometrica* 75(5):1501-1511.
- Frey, Bruno S. and Stephan Meier. (2004). "Social Comparisons and Pro-social Behavior: Testing Conditional Cooperation in a Field Experiment." *American Economic Review* 94(5): 1717-1722.
- Hoel, P.G. (1971). *Introduction to Mathematical Statistics*, 4th Edition, John Wiley and Sons.
- Karlan Dean and John A. List. (2007). "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment." *American Economic Review* 97(5): 1774-1793.
- Harrison, Glenn and John A. List. (2004). "Field Experiments," *Journal of Economic Literature*, XLII (December): pp. 1013-1059.
- Landry, C., A. Lange, J.A. List, M.K. Price, and Nicholas Rupp. (2006). "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 121 (2): 747-782.
- Lawson, C.E. (2001). “Capital Fund Appeals,” in J.M. Greenfield, ed., *The Nonprofit Handbook: Fund Raising*, Third Edition. New York, Wiley.
- Levitt, Steven D. and John A. List. (2007). “What do Laboratory Experiments Measuring Social Preferences tell us about the Real World,” *Journal of Economic Perspectives*, 21 (2): 153-174.
- List, J. A. (2006). “Field Experiments: A Bridge between Lab and Naturally Occurring Data.” *Advances in Economic Analysis and Policy* 6(2), Article 2.

- 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, 215-233.
- List, J.A. and D. Rondeau. (2003) The Impact of Challenge Gifts on Charitable Giving: An Experimental Investigation. *Economics Letters* 79:153-159.
- Meier, Stephan. (2007). "Do Subsidies Increase Charitable Giving in the Long-Run? Matching Donations in a Field Experiment." working paper, Boston Federal Reserve Board.
- Randolph, W.C. (1995). Dynamic Income, Progressive Taxes, and the Timing of Charitable Contributions, *Journal of Political Economy* 103: 709-38.
- Rondeau, D., W.D Schulze and G.L. Poe. (1999). Voluntary revelation of the demand for public goods using a provision point mechanism, *Journal of Public Economics* 72, 455-470.
- Rondeau, D., G.L.Poe and W.D. Schulze. (2005). VCM or PPM? A Comparison of the Performance of Two Voluntary Public Goods Mechanisms." *Journal of Public Economics* 81:1581-92.
- Shang, Jen and Rachel Croson. (2005). "Field Experiments in Charitable Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods." Working Paper, Wharton School of Business, University of Pennsylvania.
- Vesterlund. L. (2003). The informational value of sequential fundraising, *Journal of Public Economics* 87:627-657.

Table 1 Field Experiment - Summary Results

	High Control (HC)	Challenge (CH)	Matching (MA)	Low Control (LC)
Mean	\$45.83	\$43.78	\$34.31	\$27.79
Total Raised	\$1375	\$1620	\$1235	\$945
Median	\$30	\$35	\$25	\$25
Minimum	\$5	\$10	\$15	\$5
Maximum	\$125	\$250	\$100	\$75
Std. Deviation	\$34.72	\$41.09	\$20.04	\$14.78
Observations	30	37	36	34
Undeliverables	2	1	0	3
Net Response Rate	4.01%	4.94%	4.80%	4.55%

Notes: Cells represent raw data summaries for the various treatments. “Undeliverables” are those letters that were returned as undeliverable, likely because the person had moved.

Table 2
Field Experiment - Pair-wise two-tailed Student's t tests
Probability that the Mean of Individual Contributions in the Row and Column
Treatments are Equal

	Challenge	Matching	Low Control
High Control	0.829	0.115	0.012
Challenge		0.214	0.032
Matching			0.128

Table 3
Field Experiment - Pair-wise Two-tailed Generalized Rank-sum Tests
Probability that the Distribution of Contributions in the Row and Column
Treatments are Homogenous

	Challenge (CH)	Matching (MA)	Low Control (LC)
High Control (HC)	0.396	0.235	0.077
Challenge (CH)		0.241	0.038
Matching (MA)			0.128

Table 4 Experimental Design - Laboratory

Treatment	Threshold	Leadership Gift	Individual Benefit if Threshold Is Met	Excess Payoff Rate	Number of Subjects
High Control (HC)	\$45	\$0	\$9	\$0.20	48
Challenge (CH)	\$45	\$22.50	\$9	\$0.20	48
Matching (MA)	\$45	\$22.50	\$9	\$0.20	48
Low Control (LC)	\$22.50	\$0	\$9	\$0.20	48
Low Control 2 (LC2)	\$22.50	\$0	\$4.50	\$0.20	48

Table 5 Summary Results - Laboratory

	High Control (HC)	Challenge (CH)	Matching (MA)	Low Control (LC)	Low Control 2 (LC2)
Mean	\$7.53	\$6.84	\$5.76	7.33	\$5.41
Median	\$8.00	\$6.50	\$5.50	7.00	\$4.50
Minimum	\$1.00	\$0.00	\$0.00	1.00	\$0.00
Maximum	\$12.00	\$12.00	\$12.00	12.00	\$12.00
Std. Dev.	\$2.69	\$3.40	\$3.17	\$3.32	\$3.77
Observations	48	48	48	48	48

Table 6
Laboratory Experiments Pair-wise two-tailed Student's t tests
Probability that the Mean of Individual Contributions in the Row and Column
Treatments are Equal

	Challenge (CH)	Matching (MA)	Low Control (LC)	Low Control 2 (LC2)
High Control (HC)	0.271	0.004	0.747	0.002
Challenge (CH)		0.111	0.474	0.055
Matching (MA)			0.020	0.625
Low Control (LC)				0.010