

Do Beliefs About Peers Matter for Donation Matching? Experiments in the Field and Laboratory*

Laura K. Gee¹ and Michael J. Schreck²

¹Tufts University

²Analysis Group

August 31, 2016

Working Paper

*Gee: Department of Economics, Tufts University, Medford, MA 02155, email laura.gee@tufts.edu. Schreck: Analysis Group, Boston, MA 02199, email schreck@email.virginia.edu. We thank the John Templeton Foundation and the Science of Philanthropy Initiative for financial support. This research was approved by the IRBs of Tufts University and the University of Virginia, where Schreck completed doctoral studies. We would also like to thank Charles Holt, Federico Ciliberto, Leora Friedberg, Nathan Larson, Molly Lipscomb, Jonathan Meer, Mike Norton, Sarah Smith, workshop participants at the University of Virginia, University of Michigan, University of San Francisco, Tufts University, and University of California: San Diego, and participants at the Economic Science Association and Science of Philanthropy Initiative Conferences for their helpful comments. We thank Darcy Covert, Mike Manzi, Eitan Scheinthal, Tom Tagliaferro, and Isabelle Vrod for excellent research assistance.

Abstract

Charitable giving has been about 2% of US GDP since the turn of the century. A popular fundraising tool is donation matching where every dollar is matched by a third party. But field experiments find that matching does not always increase donations. This may occur because individuals believe that peer donors will exhaust the matching funds. We develop a theory of how beliefs about peers' donations affect one's own likelihood of donation. We test our theory using novel "threshold match" treatments in field and laboratory experiments. These treatments form small groups and offer a flat matching bonus if a threshold number of donations is received. One "threshold match" treatment more than doubles the donation rate in the field relative to no match. To better understand the mechanism behind this huge increase, we use a lab study to replicate the field results and further show that beliefs about peers' donations matter. Our theoretical, lab, and field results combined suggest people are more likely to donate when they believe they are more pivotal to securing matching money. Beliefs about others matter, and they should be taken into account when trying to increase donations.

Keywords: Charitable Giving, Field Experiment, Beliefs, Public Goods

JEL Codes: C93, D64, H41

1 Introduction

US charitable contributions totaled \$335 billion in 2013, with the average US household giving \$2,974 (GivingUSA, 2013). Charitable giving has been about 2% of US GDP since the turn of the century (Karlan, List, and Shafir, 2011). Fundraisers have begun to embrace the use of randomized field experiments to identify the best practices for attracting donors. A popular fundraising institution is to offer matching funds (e.g., an anonymous third party will match each dollar given with an additional dollar up to \$10,000). Recent field experiments find that donation matching yields varied success in terms of response rate and donation amounts. Some find that the match improves one or both of these outcomes, while others find no effect or even negative impacts.¹ This paper uses theory coupled with field and lab experiments to explain these inconsistent results. We explore how individuals' beliefs about the actions of others (henceforth, "beliefs about peers") relate to donations. Specifically, we find that people are more likely to donate when they believe they are more pivotal to securing matching money.

To our knowledge, there is no empirical work that examines the importance of beliefs about peers in the donation matching context.² Yet it is highly plausible that beliefs about peers would impact the effectiveness of the match. In non-matching contexts, previous evidence suggests that beliefs about others' contributions are correlated with own contributions in the provision of public goods in the lab and field (Croson, 2007; Fischbacher and Gächter, 2010; McBride, 2010; Fehr and Leibbrandt, 2011; Croson and Shang, 2013), and that information about others' donations can affect own donations (Shang and Croson, 2009; Smith, Windmeijer, and Wright, 2014). While these papers do not study donation matching explicitly, we believe that their findings lend plausibility to the idea that beliefs about peers may also matter in the donation matching context.

Donation matching introduces a new channel through which beliefs about peers may matter. An individual may care about securing matching money for the charity, but may believe that others' donations will exhaust the matching money.³ In this case, the individual's

¹See, for example, Chen, Li, and MacKie-Mason (2006); Karlan and List (2007); Meier (2007); Rondeau and List (2008); Eckel and Grossman (2008); Huck and Rasul (2011); Karlan et al. (2011); Gneezy, Keenan, and Gneezy (2014); Huck, Rasul, and Shephard (2015).

²This is with good reason. First, beliefs are not observable, and it is unnatural and potentially infeasible to elicit beliefs from individuals in the field. Second, in empirical settings it is difficult to disentangle the effect of beliefs about peers from other factors that may be driving contribution behavior. For example, it is plausible that a matching commitment serves as a signal that enhances the credibility of the charity to individuals (Vesterlund, 2003). In this case, the matching treatment would potentially confound the role of individuals' beliefs about peers with the role of signaling.

³Another pathway could be that the match could reduce the marginal utility derived from the public good by increasing the amount of the public good, thus reducing donations. This outcome is more likely when the individual believes that peer giving will be high, because the match multiplies the level of peer

low probability belief that her donation will secure matching money could make her unlikely to donate. The belief that one’s donation may not actually be matched could explain why traditional donation matching has inconsistent effects. This line of thinking suggests that if the match could be structured so as to *increase* the individual’s belief she is pivotal to securing matching money, then she would be more likely to give. Pivotalness is known to matter for contributions to threshold public goods in laboratory experiments (Chen, Au, and Komorita, 1996; McBride, 2010). Here we ask whether this insight can be leveraged to increase donations in the field.⁴

Using experiments in the field and laboratory, we empirically investigate the hypothesis that individuals will be more likely to donate as their belief of being pivotal to securing matching money rises. We use a novel donation matching procedure, where individuals are placed in groups of 10 and offered a (flat) matching amount of \$50 if a threshold number of donations is received (e.g., if at least 3 people donate, then the charity receives \$50 extra).⁵ We call this matching procedure the “threshold match.”⁶ In both the field and laboratory we induce variation in individuals’ beliefs of being pivotal by varying the threshold required to secure the matching money.⁷

In our field experiment, we conduct a direct mail campaign targeted toward alumni of a non-profit educational program. We use a between subjects design. We compare the performance of the threshold match treatments, with thresholds of 1, 2, and 3 donors, to a control and a traditional dollar-for-dollar match resembling Rondeau and List (2008).

giving (Karlan and List, 2007).

⁴We note that important differences exist between public goods games in the lab and charitable giving decisions in the field, including the decision-making environment and the beneficiaries of contributions. Moreover, unlike donation matching funds, linear public goods funds in the lab generally cannot be exhausted.

⁵In related work Charness and Holder (2016) find that forming groups that compete for matching funds increases the amount of contributions in a lab experiment.

⁶This threshold match bears similarities to threshold public goods games. In a threshold public goods game the public good is provided if and only if total donations (or the number of donors) reach a certain level. See, for example, Isaac, Schmidtz, and Walker (1989); McBride (2010); Anik, Norton, and Ariely (2014). Note that when a threshold public goods game calls for a binary decision (or minimum contributing set) (Van de Kragt, Orbell, and Dawes, 1983; Rapoport and Eshed-Levy, 1989), then obtaining the threshold depends only on the number of subjects who contribute. A special case is the volunteer’s dilemma, where a public good is provided if exactly one person volunteers (Diekmann, 1985). The main difference between our “threshold match” and either a threshold public goods game or a volunteer’s dilemma is that when the threshold number of individuals is obtained a bonus is paid directly to the charity rather than directly to the individual.

⁷Previous work in the laboratory and field has found that as the threshold rises there is first an increase and then a decrease in the likelihood and amount of contributions (Isaac et al., 1989; Anik et al., 2014). This observed behavior is consistent with the beliefs mechanism we implicitly test in the field and explicitly test in the lab. For example, a person may believe she is not pivotal at extremely low or high thresholds because at low thresholds others will surely donate enough, and at high thresholds the goal is extremely unlikely to be met. Indeed, we find in the lab that individuals’ elicited beliefs of being pivotal vary by the level of the threshold.

Relative to control, we find an economically substantial and statistically significant effect of the “3 out of 10” threshold match treatment. Specifically, the donation rate (3.7%) is more than twice that of the control (1.6%). A model that incorporates beliefs about peers has explanatory power.

To better trace out the beliefs mechanism, we conduct a within subjects laboratory experiment. Our laboratory experiment replicates the incentives of the field experiment while also eliciting individuals’ beliefs about peers. We view our field experiment as identifying a treatment effect of practical interest, and our laboratory experiment as testing the mechanism underlying the effect of interest and the robustness of the treatment effect.⁸ In the laboratory, subjects are still in groups of 10 with a matching bonus of \$50. But in the lab we elicit beliefs about peers and test a wider range of threshold matches, with thresholds of 1, 3, 5, 7, and 10. Our laboratory results provide further support for the beliefs explanation, as higher donation rates are associated with higher elicited probabilities of being pivotal to securing the match.⁹ Moreover, the treatment effects we observe in the field are replicated in the laboratory. In particular, the 3 out of 10 threshold match treatment has the highest donation rate, and it more than doubles the donation rate of the control.

We view the primary contributions of our work as follows. First, we provide a theoretical framework that illustrates the role of beliefs about peers in both the threshold match of our experiment and the traditional dollar-for-dollar match.¹⁰ Second, we demonstrate that this explanation is consistent with empirical results from experiments that we conduct in the field and laboratory. We find empirical support that beliefs about peers are an important factor in the decision to give when a donation match is available. This is consistent with the notion that beliefs about peers may explain the varied success of traditional dollar-for-dollar matches. Third, and of particular importance to practitioners, we document a threshold match treatment effect that is large relative to previous matching literature. Notably, our most effective threshold match treatment can be used even with relatively little matching money from a third party. Our results indicate that charities may improve donation outcomes by using matching money in more innovative ways than a traditional dollar-for-dollar

⁸See Kessler (2013) for a similar pairing of experiments in the laboratory and field.

⁹The most closely related paper to our work is McBride (2010), which finds that in the context of a laboratory threshold public goods game, the likelihood of contribution rises with the belief of being pivotal. Our work differs from McBride (2010) in two important ways. First, we combine analyses from both the field and the lab. Second, when the threshold is met in our setting, the matching money goes to the charity, rather than being paid out to the subjects as in a threshold public goods game.

¹⁰Both types of matches nest within our general framework. In our framework each individual can contribute an amount of money. And their contribution may secure matching money, but whether it does depends on the actions of peers. And the individual may hold beliefs about peer actions that affect their own decision to give.

match.¹¹

The remainder of this paper proceeds as follows. Section 2 develops a motivating theoretical model that shows how contribution behavior may depend on individual beliefs about peers when a match is available. Section 3 discusses the design and results of the field experiment, while Section 4 discusses the design and results of the laboratory experiment. Section 5 describes how alternative explanations fail to explain our results, and Section 6 concludes.

2 Theoretical Model

This section presents a general theoretical model for the decision to give charitably when a donation match is available. With simplifying assumptions, the model illustrates how our experimental treatments permit inference about individual beliefs.

Begin by letting N denote the number of agents (potential donors) in a group. Each agent is endowed with wealth w_i , which can be split between charitable contributions b_i or consumed privately as numeraire y_i . Denote the sum of charitable contributions across the N agents as $B = \sum_{i=1}^N b_i$. A donation match M is available and is a function of the vector of individual contributions: $M = \psi(b_1, b_2, \dots, b_N)$. Total contributions G is equal to the sum of individual contributions B and the donation match M .

Consistent with Andreoni’s impure altruism model (Andreoni, 1990), agent i may derive utility from consumption of the numeraire y_i , from the charitable contribution b_i (“warm glow”), and from the total contributions G . To allow for a more general model, utility flows over the sum of all contributions B and/or the donation match amount M can be considered.¹²

With choice sets and preferences established, the next step is to characterize optimal behavior for individual agents given their beliefs about other agents.

One possibility for the final step of closing the model involves invoking assumptions of rational expectations, for which each agent’s beliefs about the actions of other agents turn out to be accurate. Intuitively, the plausibility of accuracy of beliefs may vary by context. For example, in one-shot games for which agents have not accumulated information about

¹¹In the field experiment we find that a traditional dollar-for-dollar match has a donation rate of 2.34%, while our 3 out of 10 threshold match has a donation rate of 3.68%. While this is a large increase in terms of economic significance, estimate imprecision precludes a finding of statistical significance.

¹²Utility flows over these dimensions are less established in the literature, but still plausible. It is not clear that individuals view donations from peers as a perfect substitute for donation matching money that goes to the charity - a condition that is implied by specifying utility over G but not M or B . For example, Karlan and List (2007) suggest that individuals may perceive the matching third party donor as fictional. Taken literally, this suggests a utility function that contains B - the total contributions excluding the matching third party donor - but not G or M .

their peers, the rational expectations assumption may be less plausible, and may be relaxed in favor of an alternative assumption that closes the model.¹³

We note that both the threshold match of our experiments and a traditional dollar-for-dollar match nest within this organizing framework. Under both types of matches (i) each individual can contribute an amount of money; (ii) that contribution may secure matching money, but whether it does depends on the actions of peers; and (iii) the individual may hold beliefs about peer actions that affect her own decision to give. Within our framework, the difference between these two matching schemes is operationalized through the function $\psi(b_1, b_2, \dots, b_N)$, which maps individual contributions into a matching amount.¹⁴

We apply simplifying assumptions to the general theoretical framework described above to generate a model of individual optimization motivating the idea that individual beliefs about peers can play an important role in the decision to donate.¹⁵ A key component of the model is a specified threshold for donations that, if achieved, produces an amount of matching money given to the charity. Under certain assumptions, variation in the threshold permits inference about individual beliefs. This corresponds directly to the design of threshold match treatments in our lab and field experiments.

We specify a particular functional form for how the vector of contribution decisions (b_1, b_2, \dots, b_N) maps to a corresponding donation match M . Let T be an integer between 1 and N (inclusive) that represents a threshold. If at least T agents in the group choose to donate, then a fixed matching amount of b_M is provided to the charity by a third party. Because this particular functional form only requires knowing whether each agent donated (and not the amount), we restrict the giving decision to be binary.¹⁶ Specifically, the agent

¹³Relaxing the rational expectations assumption that beliefs coincide with actions is a key part of level-K models (Crawford and Iriberri, 2007) and the noisy introspection model of Goeree and Holt (2004).

¹⁴These different matching functions correspond to one conceptual difference between the threshold match and traditional dollar-for-dollar matches. It is customary to think of dollar-for-dollar matches in terms of match rate (each dollar contributed will be matched at a one-for-one rate, so long as matching funds have not been exhausted). With the threshold match, there is no match along this intensive margin. Rather, the match is secured along the extensive margin. Additionally, it is important to note that the ex-post ratio of obtained matching money to non-match donations does not represent an ex-ante match rate. For example, suppose that the threshold is 3 out of 10 and the matching amount is \$50. Suppose further that each of five donors gives \$5. As a result, \$50 of matching money were obtained from \$25 of non-match donations. This does not imply a matching rate of 2:1; it is merely a ratio of observed outcomes.

¹⁵We present a model without rational expectations here because the empirical evidence from both the lab and field indicates that subject beliefs about peers are generally not consistent with peer actions. In the lab only, 30% of the sample have beliefs that correctly guess the true realized number of givers (allowing for an error of plus or minus 1 person). We do not elicit beliefs about peers in the field, but donation rates being below 4% in every treatment implies that the individual is most likely to be pivotal in the threshold 1 (T1) treatment. But this treatment actually has the lowest donation rate of any treatment with matching money, suggesting that individual beliefs are not consistent with peer actions. See Goeree and Holt (2005) and Myatt and Wallace (2008) for a rational expectations quantal response equilibrium model.

¹⁶The binary decision simplifies the intuition of the model while also aligning it more closely with the

may choose to set b_i equal to a fixed contribution amount \bar{b} , or to 0 by not contributing.

From the perspective of the individual, the number of others who will choose to contribute is a random variable; denote this by k . The individual has a subjective belief distribution regarding the value of k that will be realized. Parameterize this distribution by letting p_j denote the subjective probability that the realized number of donors k is equal to exactly j donors, where j ranges from 0 to $N - 1$. These subjective probabilities must sum to one: $\sum_{j=0}^{N-1} p_j = 1$.

The individual agent derives utility from three things: consumption of the numeraire good, warm glow from giving (Andreoni, 1990), and matching money obtained by the charity.¹⁷ Numeraire consumption is represented by linear utility. Warm glow from giving is represented by the term $\gamma_i f(\bar{b})$ if the agent decides to give, and 0 otherwise. Matching money secured by the charity yields a utility flow of $\theta g(b_M)$ if the threshold is actually met, and 0 otherwise.¹⁸ With additional assumptions of additively separable utility flows and risk-neutral preferences, the decision to donate, characterized by comparing the expected utility of contributing and the expected utility of keeping, is described by the two expressions listed below.¹⁹

The expected utility of contributing is given by:

$$\begin{aligned} EU_{contribute} &= Pr(k \geq T - 1) * [(w_i - \bar{b}) + \gamma_i f(\bar{b}) + \theta g(b_M)] \\ &\quad + Pr(k < T - 1) * [(w_i - \bar{b}) + \gamma_i f(\bar{b})] \\ &= w_i - \bar{b} + \gamma_i f(\bar{b}) + \sum_{j=T-1}^{N-1} p_j * \theta g(b_M) \end{aligned}$$

empirical results of our experiments, which focus on the binary decision of whether to contribute. Additionally, certain matching procedures currently used by practitioners take the form of “If we get T more donors” rather than conditioning the match on attaining a given amount of money.

¹⁷Note that in this model, there is no utility derived from the total level of the public good. This is for two reasons. First, including this factor would complicate the intuition of the model but not change the conclusion that beliefs about peers are important (in fact, it would make this point even more strongly). To see this, note that one would have to integrate over the belief distribution regarding the amounts contributed by others. This would add terms to the inequality governing the decision to give, but not remove the term that includes p_{T-1} . The second reason is based on suggestive evidence from the contingent valuation method (CVM) literature that public good contributors are insensitive to the scope of the public good provided; see Karlan and List (2007) for a discussion.

¹⁸In this binary decision model, these utility flows could simply be parameterized as constants. However, this notation is chosen to maintain consistency with the model of Landry, Lange, List, Price, and Rupp (2006), which involves a continuous decision of what amount to contribute.

¹⁹We assume risk neutrality here for expositional convenience, but our result holds for all risk types. See the Appendix for a proof. We do not address risk as it relates to excuses for giving behavior (Exley, 2015).

while the expected utility of keeping is given by:

$$\begin{aligned} EU_{keep} &= Pr(k \geq T) * [w_i + \theta g(b_M)] + Pr(k < T) * [w_i] \\ &= w_i + \sum_{j=T}^{N-1} p_j * \theta g(b_M) \end{aligned}$$

A decision to contribute occurs if $EU_{contribute} > EU_{keep}$, which reduces to:

$$p_{T-1} * \theta g(b_M) + \gamma_i f(\bar{b}) > \bar{b}$$

The right-hand side of the inequality is interpreted as the certain loss in utility from contributing, since a contributor loses \bar{b} units of numeraire consumption. The left-hand side of the inequality is interpreted as the expected gain in utility from contributing. With probability p_{T-1} , the contribution is pivotal and results in a positive utility flow of $\theta g(b_M)$. A contributor also realizes a (certain) warm glow utility flow of $\gamma_i f(\bar{b})$, which guarantees a decision to contribute if it exceeds \bar{b} . Because a higher value of p_{T-1} makes the inequality more likely to hold, individual beliefs about the behavior of peers play an important role in the decision to contribute. In other words, the stronger the belief that peer contributions will sum to just below the threshold, the greater the likelihood of contributing.

Supposing now that subjective beliefs are single-peaked, we consider a comparative static on the probability of being pivotal by moving the threshold T . However, deriving this comparative static requires a modeling decision regarding the effect of the threshold level on the subjective belief distribution. To illustrate this point, consider the (potentially objectionable) assumption that this effect is negligible.²⁰ It is helpful to consider this case graphically.

Supposing that $T - 1$ is on the left-side of the peak, consider a small increase in T . Because the probability mass function is increasing in this region, there is now a greater probability that the individual's donation will be pivotal. Thus, such a threshold increase results in a greater likelihood of contributing. Consider instead an increase in the threshold T given that $T - 1$ was originally located to the right-side of the peak. By similar reasoning

²⁰This would be achieved in a model in which the individual believes that others ignore the threshold in making their contribution decisions. Beliefs that others are “naive” in this way is somewhat analogous to the beliefs of level-1 types in a level-k model, which involve other agents making decisions at random and thus being insensitive to parameter changes (Stahl and Wilson, 1994; Nagel, 1995).

as above, this increase in the threshold reduces the likelihood of contributing. Thus, in the context of this model, observing contribution decisions while holding other factors constant and varying T permits inference about the subjective belief distribution of the agent.

Suppose instead that moving the threshold T results in a different subjective belief distribution. In this case, varying T introduces new probability parameters rather than providing additional information about a single fixed belief distribution. Still, an inference can be made by comparing donation rates across the different values of T . That is, if a threshold increase results in a higher donation rate, it can be inferred that individuals hold a stronger belief that other contributions will sum to just below the threshold. This is the intuition behind the design of the various threshold-match treatments in both the field and lab experiments (T1, T3, etc.).

3 Field Experiment

The field experiment was conducted in cooperation with a 501(c)3 charitable organization in the United States that operates educational programs for students. This organization traditionally conducts two fundraising appeals per year using direct mail through the United States Postal Service. The first appeal typically occurs in early summer, while the second occurs in early winter. This experiment was conducted during the 2012 summer fundraising window.

The sample of subjects includes 2,567 individuals who attended a program for high-achieving high school students that is operated annually by the organization. This program, which produces approximately 100 alumni each year, attracts students from a large geographic area encompassing many states. The sample was constructed by including all living individuals who completed the program between the years 1977 and 2007, omitting those individuals who either had no valid mailing address or had opted out of receiving mailings.

3.1 Field Experiment Treatments

In a between-subjects design, individuals were randomly assigned to one of five treatments. Treatments varied language in the fundraising letter about the extent to which an individual's donation would impact the amount of matching money given to the charity. Aside from one paragraph (approximately two-thirds of the way down the letter) that varied across treatments, the letter was the same for all subjects.

The control (C) treatment included standard fundraising language that was similar to language used in previous appeals. The matching (M) treatment offered a dollar-for-dollar

match up to \$3,000, noting that the individual had been selected for a group of 600 alumni eligible for the match. The threshold-match (T1, T2, T3) treatments notified the individual that they had been selected for a group of 10 alumni who attended the program in their attendance year, and if a threshold number of these individuals chose to donate, then a matching third party would provide a fixed amount of \$50 to the organization. The T1 treatment involved a threshold of 1, while the T2 and T3 treatments featured thresholds of 2 and 3, respectively.²¹ Note that the total amount of \$3,000 for 600 people chosen in the traditional matching treatment (M) is equivalent to a possible \$50 bonus for each group of 10 people in this treatment. A generous benefactor had agreed to provide matching funds in accordance with the rules stated in the letters, thus ensuring that no individuals were deceived.

The language for the various treatments is presented below, with italics, bolding, and underlining accurately replicated:

- Control (C): “Your decision to give can make an *important, positive impact* on [organization name]. The **generosity of alumni like you helps ensure that talented young [11-word description of student population served] are able to experience our unique [2 adjectives describing program] education programs.** To help support us in these endeavors, please submit your donation by August 31.”
- Matching (M): “Your decision to give can make an *important, positive impact* on [organization name] - and it can go even further thanks to a commitment of matching donations that we have received. As part of this particular donation program, you are one of 600 randomly selected alumni eligible for a **dollar-for-dollar match**, up to a total group limit of \$3,000. This means that **for each of the first \$3,000 we receive from your group of 600, we will also receive a matching dollar of support, thus doubling the impact of your donation.** To take advantage of this matching opportunity, please submit your donation by August 31.”
- Threshold Match (T1, T2, T3): “Your decision to give can make an *important, positive impact* on [organization name] – and it can go even further thanks to a commitment of matching donations that we have received. As part of this particular donation program,

²¹Some may worry that subjects in the threshold-match treatments might communicate with each other - unobserved to researchers - in an effort to coordinate on reaching the threshold. We think this implausible for several reasons. First, subjects assigned to a common threshold-match group were anonymous to each other. Second, given the passage of time since subjects attended the program and the large geographic area from which subjects are drawn, it is highly unlikely that subjects would attempt to or be successful in unraveling this anonymity. Finally, notwithstanding the success of the T3 treatment in terms of increased donation rates, no group with a threshold of 2 or 3 was actually successful in reaching the threshold.

you have been randomly and anonymously selected for a group of 10 [students] who attended [program name] in your year. **From your group of 10, if at least (1 person chooses; 2 people choose; 3 people choose) to make a donation of any amount, then the matching donor will provide an additional \$50 of support, thus increasing the impact of your donation.** To take advantage of this matching opportunity, please submit your donation by August 31.”

3.2 Field Experiment Randomization and Constraints

In order to prevent deception and to improve the ability to make inference, certain constraints were incorporated into the randomization routine that assigned treatments. Recall that the threshold-match letters informed individuals that they had been randomly and anonymously selected for a group of 10 students from their attendance year. To prevent deception, the randomization was programmed to meet this constraint.²² A further concern was achieving balance across treatment cells in previous giving history, which would be expected to predict future giving behavior (Meer, 2013).²³ We balanced on an indicator variable (warm4) that takes the value of 1 if the individual had donated at least once in the 4 years prior to the start of the experiment and 0 otherwise. The balance of this variable across treatments was also programmed to hold within attendance years at the program, to reduce the correlation of age and cohort variables to treatment. Finally, the amount of available matching funds was limited, so it was not feasible to assign equal numbers of individuals to each treatment. Balance was achieved on all pre-experiment attributes with the exception of cohort year, which was slightly higher in the Control, but which we correct for in our regression analysis with the inclusion of cohort level fixed effects. A balance table on pre-experiment attributes is available in the Appendix.²⁴

²²Ex-ante the randomization always truly assigned a group of 10 people from each cohort attendance year for T1/T2/T3 and 600 people for the dollar-for-dollar match (M). However, a few letters were returned, so that ex-post some groups had fewer than the originally reported number of people. People attended between 1977 and 2007 and the cohort sizes are between 70 and 101 people.

²³While ex-post adjustments can be made for imbalance, it is more efficient to achieve ex-ante balance (Bruhn and McKenzie, 2009).

²⁴One may worry that subjects in the experiment may have colluded with each other. This is unlikely, as the subjects attended the program between 1977 and 2007 and currently are very geographically spread out (residing in over 81 different countries/states). Also the timing of donations is very similar across all the treatments (results available from authors upon request).

3.3 Field Experiment Results

Table 1 summarizes the response to the fundraising appeal within the fundraising window, which lasted just over two months.²⁵ The first row of Table 1 shows that donation rates vary across treatments, in a range from approximately 1.6% to 3.7%. For many fundraisers the donation rate is of primary importance. For example, universities are quite concerned with the alumni donation rate because this is used by US News and World Report for their annual college rankings.²⁶ The number of donations is low enough (relative to the variation in individual donation amounts) that the mean donation conditional on giving is sensitive to outliers. To better characterize the response distribution, the 20th, 50th, and 80th percentiles (conditional on giving) are provided. These data suggest that the response distributions of the various treatments exhibit significant overlap, as the median is the same across all treatments. Indeed, this intuition is formalized by the results of a non-parametric Kruskal-Wallis equality-of-population test. Neither this test nor pairwise Wilcoxon rank-sum tests can reject the null hypothesis that the amount distribution conditional on giving is the same across treatments. This result suggests that individuals do not reduce donation amounts as a result of the threshold match treatment, which is interesting because the match provides no explicit incentive to donate along the intensive margin (i.e., they could donate only a dollar and still count toward securing the match).²⁷

²⁵Use of this outcome measure maintains consistency with Karlan and List (2007); Karlan et al. (2011). Donations received after the fundraising window also comprise important information. We observe no meaningful change in outcomes when the end of the window is extended by one month.

²⁶See <http://www.usnews.com/education/best-colleges/articles/ranking-criteria-and-weights>.

²⁷However, this does not rule out the possibility that our treatments have a positive effect on the amount donated conditional on donating. For example, it is possible that a treatment increases the amount donated conditional on donating for one type of individual, but also increases the propensity to donate for another type of individual. If this second type is inclined to give small amounts, this counteracts the positive impact on amount donated from the first type, and the observed donation amount distribution could remain unchanged. For more discussion of this selection effect, see Karlan and List (2007).

Table 1: Field Experiment: Summary Statistics by Treatment

Variables of Interest	Treatments				
	C	M	T1	T2	T3
Donation rate (0 to 100)	1.59	2.34	2.00	2.35	3.68
Mean Donation, conditional on giving	255.59	158.93	195.83	94.28	158.45
Dollars raised per letter, excluding match	4.06	3.71	3.92	2.21	11.79
Dollars raised per letter, including match	4.06	7.43	4.75	2.21	11.79
Percent of Potential Third Party Match Funds Obtained (0 to 100)	NA	74.2	16.7	0	0
Observations	1,071	599	300	298	299
Amount Percentile 20, conditional on giving	50	50	100	50	75
Amount Median, conditional on giving	100	100	100	100	100
Amount Percentile 80, conditional on giving	350	250	150	250	250

Note: The Control (C) treatment received a standard letter. The Match (M) treatment received 1-1 matching. The threshold match treatments (T1, T2, and T3) involved groups of 10 people and if the threshold number of persons in a group gave (1, 2, or 3 respectively) then an extra \$50 would be given to the charity.

With small donation counts, amount outliers, and overlapping amount distributions, there is little hope of precisely estimating any type of treatment effect on donation amount. Thus, we focus on the binary decision of whether one contributes. Table 2 presents least squares regressions with the binary outcome variable of whether the individual contributed during the fundraising window. The first column presents a simple linear model with dummy variables for the treatments and a control for previous giving history (*warm4* which takes the value 1 if a person donated in the past four years and 0 otherwise). The second column presents the same simple model but clusters standard errors at the cohort level. The third column adds in cohort level fixed effects. Last, the fourth column weights observations by the inverse of the sampling probability.²⁸ Across the four specifications, the only treatment with a statistically significant effect on donation rates as compared to the control was the Threshold Match 3 (T3) treatment. The coefficients on T3 range between 1.945 to 2.008 percentage points, illustrating that the Threshold Match 3 (T3) treatment more than doubled the donation rate as compared with the mean of 1.59% in the control group. These results are significant at the 5-10% level.²⁹ In terms of percent change, the T3 treatment effect (131

²⁸The average cohort year had 86 observations (min: 70; max: 101). In each cohort year, 20 alumni were assigned to M, 10 were assigned to each treatment of T1/T2/T3, and the remaining alumni were assigned to C. Thus in years with more observations an individual is more likely to be assigned to the control. For example if the 1997 cohort had 80 individuals then probability of being assigned to T1/T2/T3 is $\frac{10}{80}$, M is $\frac{20}{80}$, and control is $\frac{30}{80}$. Whereas if the 2004 cohort had 100 individuals then probability of being assigned to T1/T2/T3 is $\frac{10}{100}$, M is $\frac{20}{100}$, and control is $\frac{50}{100}$. We adjust for this sampling by presenting results which have been weighted by the inverse of the sampling probability alongside the un-weighted results.

²⁹Results from a probit specification are similar. Results using other indicator variables of previous giving in the last year (*warm1*), 2 years (*warm2*), 3 years (*warm3*), or five years (*warm5*) are also similar and are presented in the Appendix.

percent increase) is quite large relative to extensive margin treatment effects observed in the donation matching literature, which tend to range from slightly negative values to as large as 31%.³⁰

Table 2: Field Experiment, Dependent Variable: Probability of Donation (0 to 100)

	Simple 1	Simple + Cluster SE 2	Cohort FE 3	Cohort FE + Weighting 4
Match (M)	0.850 (0.697)	0.850 (0.809)	0.793 (0.819)	0.731 (0.829)
Threshold Match 1 (T1)	0.438 (0.893)	0.438 (0.976)	0.378 (0.985)	0.335 (0.975)
Threshold Match 2 (T2)	0.673 (0.895)	0.673 (0.891)	0.609 (0.902)	0.675 (0.923)
Threshold Match 3 (T3)	2.008** (0.894)	2.008* (1.040)	1.949* (1.048)	1.945* (1.051)
warm4	15.398*** (0.870)	15.398*** (2.058)	15.283*** (2.114)	15.876*** (2.694)
Adjusted R2	0.109	0.109	0.107	0.107
N	2,567	2,567	2,567	2,567
Clustered SE		X	X	X

Notes: All coefficients are multiplied by 100 for ease of readability. The variable warm4 takes the value 1 if a person donated in the past four years and 0 otherwise. Standard errors are clustered at the cohort level in columns 2-4 (there are 30 cohorts). In column 4, the observations are weighted to account for the constraints on randomization. Legend: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

None of the treatments besides T3 are found to be significantly different from control. It is worth noting, however, that estimates for each of these other treatments are all positive and potentially economically important, with donation rate increases ranging from 21% to 53% relative to the control mean. Estimate imprecision prevents statements of statistical significance at conventional levels for these treatments. Moreover, although the T3 treatment yields substantially larger donation rates than any other matching treatment (M, T1, T2), it is not statistically significantly different from these treatments. For example, the traditional dollar-for-dollar match (M) has substantially lower point estimates (between 0.731-0.850)

³⁰See, for example, Karlan and List (2007); Meier (2007); Rondeau and List (2008); Eckel and Grossman (2008); Huck and Rasul (2011); Karlan et al. (2011); Gneezy et al. (2014); Huck et al. (2015). Gneezy et al. (2014) find a non-matching treatment effect of similar magnitude to the T3 treatment effect. In their experimental treatment, a third party contributed \$10,000 to cover all overhead costs associated with raising donations for a non-profit campaign.

compared to the T3 coefficients (between 1.945-2.008), but pairwise tests do not find a statistically significant difference.

A non-profit may have a number of objectives such as increasing donation rate, amount contributed pre-match, and amount contributed post-match. Our T3 treatment has the highest donation rate, mean donation, and dollars raised per letter both excluding and including the matching funds, although as previously stated the results about amount raised per letter are sensitive to outliers. Additionally a practitioner may be concerned with the total amount of matching funds secured during a single campaign. If matching funds will be forfeited if not utilized during a single campaign, then the traditional dollar-for-dollar match (M) may be preferable since it resulted in \$3.72 of matching funds per letter.³¹ However, if only a small amount of matching funds are available or if matching funds that go unused can be utilized through another channel or at a later date, then the novel threshold matches may be particularly attractive since none of the matching funds were used in either the T3 or T2 treatments. Again, these dollar per letter amounts are sensitive to outliers, and so these results should be interpreted cautiously.

We interpret these results as suggestive, though not conclusive, evidence that structuring matching money in more innovative ways than a dollar-for-dollar match can lead to substantial improvements in donor outcomes.³²

3.4 Beliefs, Actions, and Rational Expectations

Empirical donation rates across the various treatments range from 1.6% to 3.7%. With rates this low, it is difficult to explain the effectiveness of the T3 treatment with a rational expectations model. Given the empirical donation rates, the individual is most likely to be pivotal in the T1 treatment. Note that although an individual is most likely to “actually be pivotal” in T1, that does not mean she has the highest “belief of being pivotal” in T1. In the field we cannot directly observe beliefs, which is exactly why we have coupled our field experiment with a lab experiment. Donation rates were monotonically increasing as the donation threshold was raised. These results, together with the intuitive idea that individuals likely do not have precise knowledge about the donation behavior of their peers, suggest that to accurately understand giving outcomes in this context, the assumption of beliefs being consistent with actions should be relaxed. This contrasts with a number of models of giving behavior in the literature that impose rational expectations to draw conclusions (e.g.,

³¹We note, however, that only 74.2% of the available dollar-for-dollar match was secured in our field experiment.

³²We note that the large effect from the innovative overhead treatment in Gneezy et al. (2014) is consistent with our proposition, with the caveat that the overhead sponsorship is an unconditional gift, whereas matching money is a conditional gift.

Andreoni (1989, 1990); Landry et al. (2006); Karlan and List (2007)).

4 Laboratory Experiment

A novel explanation involving beliefs about donations by peers is successful in explaining the treatment effect in the field experiment. However, there are two shortcomings of the field experiment that prevent a cleaner test of the beliefs story. First, individual beliefs about peers are not observed or elicited in the field. Second, the donation threshold varies across a small range, from 1 to 3 out of 10 individuals.

These drawbacks motivate a laboratory experiment that replicates the incentives of the field experiment but allows resolution of the shortcomings discussed above. To address the first concern, this experiment elicits individuals' beliefs about their peers. This provides a more controlled, supplemental test of the explanation that beliefs about peers are an important factor in the decision to give charitably. To address the second concern, the laboratory experiment extends the range of thresholds by considering thresholds of 1, 3, 5, 7, and 10 donors. With five different thresholds spread almost uniformly throughout this range, one can consider the experimental treatments as tracing out portions of the threshold-to-donation-rate relationship that the field experiment does not explore.³³ This experiment also provides a robustness check of two results from the field experiment: (1) substantial variation in donation rate across thresholds, and (2) the superior performance of the 3 out of 10 treatment relative to control.

4.1 Laboratory Design and Procedures

Between the field experiment discussed earlier and this laboratory experiment, many of the general decision properties remain, such as the number of people in each group, the charitable giving nature of each person's decision, and certain treatments governing the rules for how individual decisions translate into matching money that will be given to the charity. However, a number of variables change between the two experiments, such as the subject pool, charity, method of solicitation, between vs. within subject design, and the physical location where the donation decision is made.

Subjects in this laboratory experiment were recruited from the student population at Tufts University. At the beginning of each session, the experiment administrator announced

³³We did not explore all other possible threshold values (2, 4, 6, 8, 9) due to time constraints for each experimental session, and because we expected that they would yield limited insights beyond those offered by the threshold values that we did explore. Also, it was infeasible to replicate the field experiment's dollar-for-dollar match treatment, which involved 600 individuals.

the instructions of the experiment, then directed subjects to complete a short quiz that tested their understanding of the experiment.³⁴ Each session lasted about one hour and average earnings were approximately \$18.75. The experiment was conducted using zTree (Fischbacher, 2007).

Each session involved 10 participants and consisted of a donation stage followed by a belief elicitation stage. Each stage consisted of six periods, and each period corresponded to one of six treatments. The laboratory experiment is a within subject design, which allows us to observe variation in beliefs and donations decisions across treatments *for each subject*. This is intended to control for subject-specific effects that might correlate with beliefs but also affect the decision to donate. Additionally, because subjects see all the thresholds, there is no signaling value in the threshold level about the quality of the charity. In total, 10 sessions were conducted, thus yielding a sample size of 100 subjects. Each subject participated in each of the following six treatments:

- (C) Control: Binary decision to give \$0 or \$5, no donation match available
- (T1) Threshold Match 1: Binary decision to give \$0 or \$5, with a flat \$50 matching donation if at least 1 out of the 10 subjects donated
- (T3) Threshold Match 3: Identical to T1 but with a threshold of 3 subject donations
- (T5) Threshold Match 5: Identical to T1 but with a threshold of 5 subject donations
- (T7) Threshold Match 7: Identical to T1 but with a threshold of 7 subject donations
- (T10) Threshold Match 10: Identical to T1 but with a threshold of 10 subject donations

Each treatment involved a single binary donation decision for each subject, with the non-profit organization *Citizen Schools Massachusetts* serving as the charity.³⁵ Unlike the field experiment where subjects can choose to donate any amount, we chose to limit subjects to a binary choice between giving \$0 or \$5 in the lab. First, we felt it was important to have some non-zero minimum donation in the lab experiment to better approximate the higher transaction costs of making a donation in the field experiment. In the field experiment a person must write a check, get a stamp, and physically mail the letter to the charity, whereas in the lab a person need only press a button to make a donation. Second, in the field experiment the amount raised per letter was sensitive to outliers, so by limiting subjects

³⁴Full instructions and screenshots are available in the Online Appendix.

³⁵*Citizen Schools Massachusetts* is a 501(c)3 charitable organization dedicated to expanding the learning day for middle school children in low-income communities. Like the charity in the field experiment, it is an organization focused on educational outcomes.

to give exactly \$5 we no longer allow outliers to drive the differences in amount raised per person in the lab. Last, a binary donation decision simplifies the beliefs space for subjects in the lab since they need only hold beliefs about the total number of other givers, rather than about both total number of givers and the amounts each donates.

A decision to “Donate” meant that the subject donated \$5 to the charity, and a decision to “Keep” meant that the subject donated nothing.³⁶ Donations were made from the subject’s endowment, which was \$16. Importantly, subjects knew that in the donation stage, only one period was randomly selected for payment (and donations) to occur. For the randomly selected period, subjects received payments as follows: \$16 if the subject chose “Keep” in the selected period (thus keeping their full endowment), and \$11 if the subject chose “Donate” (with the remaining \$5 going to the charity). For the threshold match treatments, subjects were informed that if the number of subjects who donated in the randomly selected period met or exceeded the threshold, then an anonymous third party would give an additional \$50 to the charity. This amount was thus consistent with the \$50 matching offer provided in the field experiment.

Across sessions, the ordering of treatments within a stage was randomly rotated, so as to avoid complications from order effects.³⁷ To prevent learning effects that might confound treatment effects, subjects did not receive any information regarding the decisions made by their peers until the conclusion of the session, after all decisions had been made.

After the donation stage had concluded, subjects began the belief elicitation stage. For each period in the donation stage, subjects were asked to report on their beliefs regarding the actions of their peers. This was done via the quadratic scoring rule (Selten, 1998) via an elicitation method used in previous experiments (Moore and Healy, 2008; Eil and Rao, 2011).³⁸ Each subject was asked to assign a probability p_j to each possible outcome of the number of other donors ($j = 0, 1, 2, \dots, 9$). The probabilities across those outcomes were required to add up to one (Manski and Neri, 2013). Depending on the accuracy of their guesses, subjects could earn up to an additional \$10 for the beliefs stage of the experiment using the following equation:

³⁶Subjects had the option of filling out a self-addressed envelope before making any decisions, so that a personal receipt for any potential donations could be mailed.

³⁷Within a given session, the same ordering of treatments was maintained between the donation and belief elicitation stages. For example, if the first period of the donation stage featured the T1 treatment, then beliefs about donation decisions in the T1 treatment were elicited in the first period of the belief elicitation stage.

³⁸The quadratic scoring rule is incentive-compatible under risk neutral preferences. Consistent with prior literature, our subjects were told that they would maximize expected earnings by telling the truth. See Holt and Smith (2013) for a lottery menu elicitation method that is incentive compatible regardless of risk preferences.

$$5 \times \left[2 - \sum_{j=0}^9 (1\{\text{actual number of peer donations} = j\} - p_j)^2 \right]$$

Only one of the elicitation stage periods was randomly selected for payment. The selection rule was as follows: omit as a candidate the period chosen for payment in the donation stage, and choose equally likely among the remaining periods. This was done to avoid the potential for subjects to hedge across their donation and belief elicitation decisions.³⁹

4.2 Laboratory Results

Figure 1 displays the donation rate by treatment. The Threshold Match 3 (T3) treatment has the highest donation rate, with half of subjects choosing to donate in this treatment. The control (C) has the lowest donation rate, with just under 25% of subjects donating. In column 1 of Table 3, we report the results from a simple linear regression predicting the likelihood of donation by treatment. Treatment T3 increases the likelihood of donating by 26 percentage points above the mean of 24% in the control (C). That represents a doubling of the likelihood of donation. In Table 3, we report both the results for the full 100 subject sample and for the sub-sample of subjects who vary their donation choices across treatments; that is, we exclude those subjects who always donated or never donated.⁴⁰ In column 2 of Table 3 we concentrate on only those subjects with variation across treatments to find that T3 increases the likelihood of donation by 42.62 percentage points over the mean of 16.39% in the control, representing a 260% increase. T3 is statistically significantly different from C

³⁹For example, suppose that the threshold is 3 and a subject (accurately) believes that their peers will give either 1 or 2 donations, with 70% probability on the 2 donation outcome. Suppose also that the subject derives utility from the threshold being achieved. Then, a risk-averse subject may hedge by choosing to donate in the donation stage (which pays off in utility terms if 2 of their peers donated), but then put disproportionate weight on the outcome 1 in the belief elicitation stage (which pays off in monetary terms if 1 of their peers donated). If both decisions are selected for payment, then the subject has guaranteed a (utility and money) payoff by hedging, but has not truthfully revealed their guess in the belief elicitation stage. By using the period-omission process described above, this incentive to hedge is eliminated. See Blanco, Engelmann, Koch, and Normann (2010) for discussion of the extent of hedging problems in belief elicitation experiments.

⁴⁰The variation in donations across treatments is fully accounted for by a proper subset of the full sample of subjects. This is demonstrated in the Appendix, which presents a histogram of the proportion of subjects who made a given number of “Donate” decisions across all 6 treatments. The subset of subjects who show variation across treatments is interesting, because it is more analogous than the full sample of laboratory subjects to the field experiment subjects who are actually responsible for driving the treatment effect. Said differently, any potential donor population can be divided into those who will never donate, those who will always donate, and those whose donation decision depends on the treatment. The final group accounts for the observed treatment effect in the field experiment, and we seek to understand whether beliefs about peers drive donation behavior in that population.

and T1 in the models in column 1 and 2, but is not statistically significantly different from T5/T7/T10.⁴¹

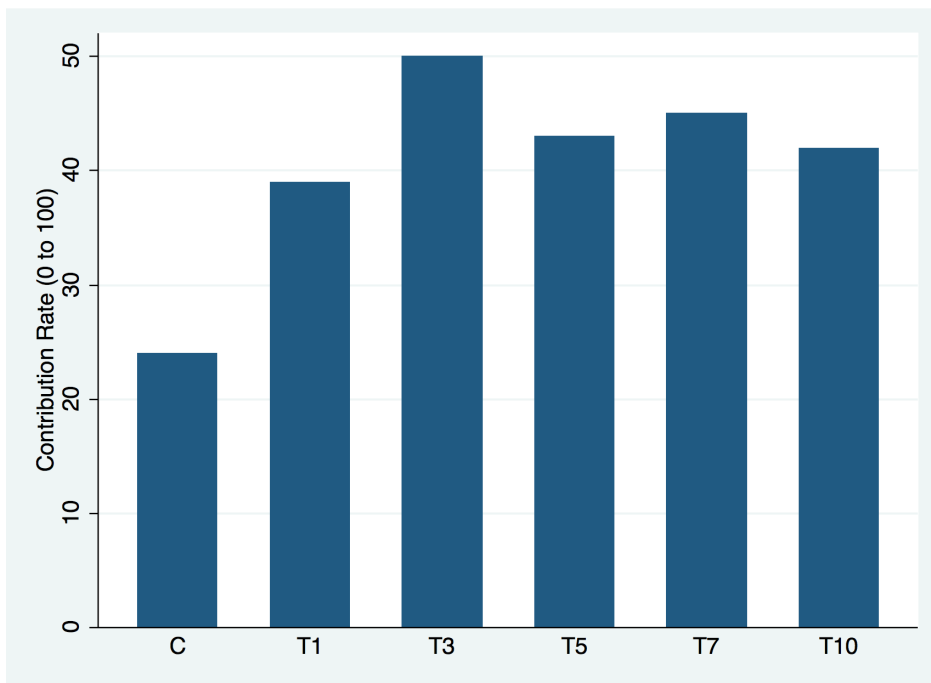


Figure 1: Lab Experiment Donation Rate by Treatment

Note: This figure shows the donation rate by treatment in the lab experiment.

Certain findings from the field experiment are robust to the changes of conducting the experiment in the laboratory. There is substantial variation in donation rate across thresholds, and Threshold Match 3 (T3) is the top-performing treatment.⁴²

We also observe an increasing monotonicity in donation rates across the control, T1, and T3 treatments. Additionally, in row three of Table 4 we see that like the field experiment

⁴¹For the column 1 specification, the test that the coefficient on T3 is equal to T1 yields $F(1, 99) = 4.16$ with $Prob > F = 0.044$, and for column 2, $F(1, 60) = 4.25$ with $Prob > F = 0.0436$. One may be worried about order effects in our experiment or that the within-subjects design leads to stronger experimenter demand effects (Charness, Gneezy, and Kuhn, 2012). First, we randomly assigned the order of the treatments, and there is no strong correlation between order and the giving decision. Second, if we restrict the analysis to only the first decision made by participants, we can show that all the treatments except T1 perform better than the control; however, these results are too noisy to say which treatment performs best. Also they do not control for time invariant individual level heterogeneity, which is exactly why we ran a within subjects experiment.

⁴²Interestingly, this result holds even though the level of donation rates is much higher in the laboratory setting than it is in the field. A threshold of 1 yields a donation rate of 2.0% in the field experiment and 39% in the laboratory experiment. Similarly, a threshold of 3 yields a donation rate of 3.7% in the field experiment and 50% in the laboratory.

Table 3: Lab Experiment, Dependent Variable: Probability of Donation (0 to 100)

	<u>With Control</u>		<u>With Beliefs</u>	
	Full 1	Variation 2	Full 3	Variation 4
Pivotal Belief Level			0.22** (0.09)	0.43*** (0.15)
T1	15.00*** (5.50)	24.59*** (8.80)		
T3	26.00*** (5.77)	42.62*** (8.73)	10.71* (5.56)	20.78** (9.59)
T5	19.00*** (5.57)	31.15*** (8.76)	3.08 (6.17)	6.22 (11.07)
T7	21.00*** (5.48)	34.43*** (8.51)	6.58 (7.21)	12.01 (12.81)
T10	18.00*** (4.79)	29.51*** (7.45)	1.71 (6.67)	2.80 (11.75)
Constant	24.00*** (3.29)	16.39*** (4.87)	36.01*** (4.35)	29.76*** (7.67)
R2	0.53	0.28	0.54	0.22
Adjusted R2	0.43	0.12	0.42	0.01
N	600	366	500	285
Subjects	100	61	100	57

Notes: All coefficients are multiplied by 100 for ease of readability. All models include subject fixed effects, and standard errors are clustered at the subject level. Legend: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

in the laboratory our T3 treatment has the highest donation rate and dollars raised per subject, both excluding and including the matching funds. In the field experiment there were relatively few donations, such that dollars raised per letter were sensitive to outliers. In contrast, in the lab subjects are forced to donate exactly \$5, so that dollars raised per subject are not sensitive to outliers. In the lab we see that dollars raised per subject, both excluding and including the match, are highest in the T3 treatment.⁴³ If a practitioner wanted to increase the donation rate, amount contributed pre-match or amount contributed post-match, both our lab and field experiments suggest that the T3 treatment would be best.

⁴³The difference between T3 and the other treatments is statistically significant when comparing pre-match dollars per subject, and is statistically significant for all but T3 vs. T1 for comparing post-match dollars per subject. All comparisons are based on comparisons of coefficients in a linear model without a constant.

Table 4: Lab Experiment: Summary Statistics by Treatment

Variables of Interest	Treatments					
	C	T1	T3	T5	T7	T10
Donation rate (0 to 100)	24	39	50	43	45	42
Mean/Median Donation, conditional on giving	5	5	5	5	5	5
Dollars raised per subject, excluding match	1.20	1.95	2.50	2.15	2.25	2.10
Dollars raised per subject, including match	1.20	6.95	7.00	5.15	3.25	2.10
Percent of Potential Third Party Match Funds Obtained (0 to 100)	NA	100	90	60	20	0
Observations	100	100	100	100	100	100

Notes: All the above amounts are based on the choice of subjects if all the rounds of the experiment had been paid. In the actual experiment only 1 randomly selected round was actually used for payment to both the subject and the charity.

The laboratory experiment explores the relationship between donation rate and treatment using threshold values larger than those used in the field experiment. While both experiments find an increasing donation rate when the threshold value is less than 3, the laboratory results demonstrate generally decreasing donation rates above the threshold value of 3. This inverted U shape is consistent with previous work that finds contributions are highest with a mid-level threshold (Anik et al., 2014; Isaac et al., 1989).⁴⁴ However the estimates are too noisy to allow any strong conclusions, as the T3 treatment is not statistically different from the T5, T7, and T10 treatments in donation rate.

The inverted U shape in Figure 1 is consistent with a beliefs explanation. The increasing portion is consistent with the story that subjects may not believe they are pivotal when there is a low threshold because they expect other donors to achieve the threshold. This amounts to a free riding explanation when the threshold is low, such as in T1. The eventual decreasing pattern as the threshold rises is also consistent with a beliefs story. That is, subjects may believe that with such demanding thresholds, their peers are not likely to donate enough to make their own contribution pivotal (e.g., with T10). Considering the increasing and decreasing portions together, the beliefs story amounts to an explanation of free riding when the threshold is low and belief that the matching money is a “pipe dream” when the threshold is high. This intuitive explanation and the results observed in the laboratory suggest that in the field experiment donation rates would eventually have declined if the threshold had been raised even further.

Our study was initially motivated by the inconsistent effects of traditional dollar-for-dollar matching on charitable contributions. We felt that these inconsistencies might be driven by beliefs about the giving behavior of others. To test if beliefs are driving heterogeneity in

⁴⁴This overall pattern is also consistent with the quantal response equilibrium theoretical model presented in Goeree and Holt (2005).

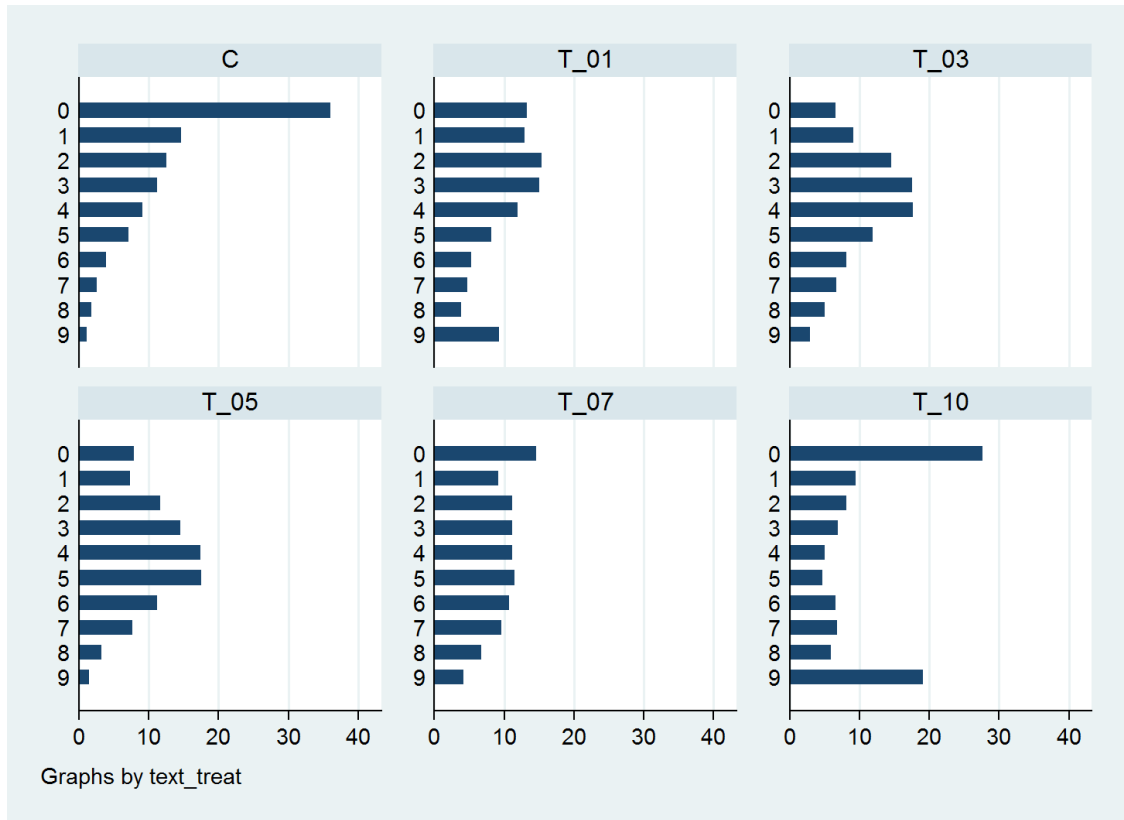


Figure 2: Average Beliefs by Treatment

Note: This figure shows the average probability placed on 0, 1, 2, ..., up to 9 other persons in the group donating. This is the average of the reported probabilities for all 100 subjects in all 6 treatments.

donation rates when matching funds are available, we need to examine the reported beliefs from the laboratory experiment. Collectively, the lab subjects tend to report single-peaked probability distributions as pictured in Figure 2.⁴⁵ Figure 2 reports the average reported beliefs of 0 up to 9 other people giving in each treatment. As the threshold moves from 1 to 5 people, there tends to be a shifting of the highest (modal) probability mass toward a larger number of givers. However, under the very high thresholds (T7 and T10) people put a large amount of mass on 0 other givers, consistent with the idea that reaching the threshold is a “pipe dream” when the threshold is very high.

On average, the lab subjects do a good job predicting the true number of realized peer donors. We can compute the expected number of peer donors as $\sum_{j=0}^9 p_j j$ and compare that to the actual realized number of peer donors in each treatment (as shown in Table 8 of the Appendix). Although collectively the realized number of peer donors is close to the expected

⁴⁵Only 6 individuals in our sample of 100 never report single-peaked belief distributions. Of the 600 total elicited belief distributions, 83% are single peaked.

number, this masks the fact that individually the subjects often report an expected number of peer donors that is 7 donors too high or 7 donors too low. In fact, individually, subjects only report an expected number of peer donors within 1 too high or 1 too low about 70% of the time.

The subjective probability of being pivotal to the match is the reported probability that $T - 1$ other subjects give in each treatment (e.g., for a threshold of 3, it is the reported probability that 2 other subjects give). Our theoretical model predicts that as p_{T-1} rises, the likelihood of donation also rises. Figure 3 shows the relationship between the probability an individual puts on being pivotal and the donation rate. The figure reveals an upward sloping pattern: higher probabilities of being pivotal trend with higher donation rates.

Figure 3 does not account for differences driven by treatment or by individual attributes (e.g., optimism). So, in columns 3 and 4 of Table 3, we report the results from linear models that include the probability that $T - 1$ others donate as an explanatory variable while including individual fixed effects and treatment fixed effects. Note that we do not randomly assign beliefs, but rather we randomly assign treatments, which influence beliefs. So we cannot say that changes in beliefs cause changes in donations, but rather that changes in beliefs are associated with different donation rates, consistent with our theoretical model. Recall that we were originally motivated to explore if beliefs about peers could explain inconsistent results in traditional dollar-for-dollar matching. Our ultimate objective is to explain the donation rate. We first show that our different threshold treatments lead to variation in donation rates in both the lab and the field. Next we show that in the lab these differences in donation rates by threshold can be explained by beliefs about peers.

A subject only has beliefs about being pivotal in our laboratory experiment if there is a threshold match, so the regression analysis in columns 3 and 4 of Table 3 only applies to the five threshold match treatments and excludes the control. In column 3, we find that a one percentage point increase in the likelihood of being pivotal (beliefs that $T - 1$ others give) is associated with a positive and statistically significant 0.22 percentage point increase in the likelihood of donation. This rises to an associated increase of 0.43 percentage points when we only concentrate on those subjects with some variation in donations across threshold match treatments (column 4). This implies that a 10 percentage point increase in belief of being pivotal would be associated with a 2.2 to 4.3 percentage point increase in the likelihood of donation.⁴⁶ As beliefs about being pivotal rise, so does the likelihood of

⁴⁶For comparison purposes we provide the same models in column 3 and 4 but without beliefs in the Appendix. Also, for comparison purposes, we describe the results for only the first period since one may be worried about order effects in our experiment. If we restrict the analysis to only the first decisions made by participants, we can show that the coefficient on pivotal beliefs is still positive but insignificant. However, these results do not control for time invariant individual level heterogeneity, which is exactly why we ran a

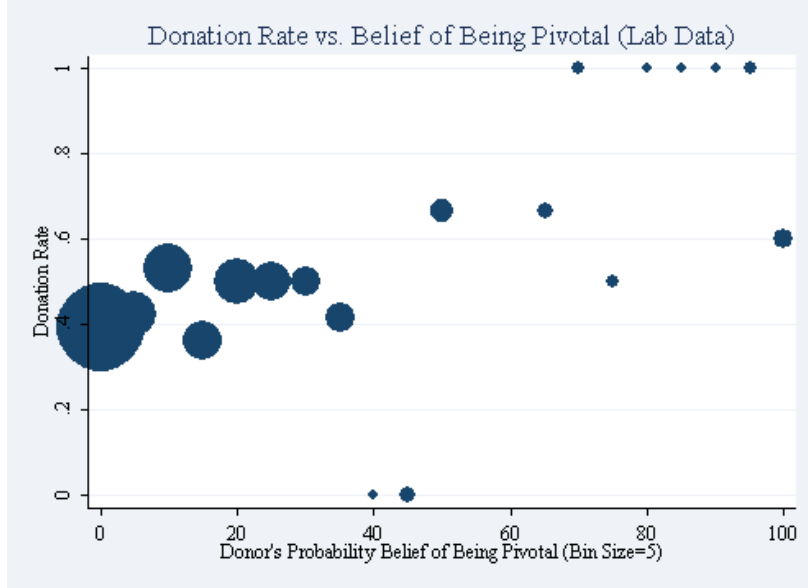


Figure 3: Donation Rate in Each Pivotal Belief Bin

Note: Sample includes only subjects ($N = 57$) who showed variation in donation decisions across treatments. Each bin has a width of 5, and the donation rate variable on the vertical axis is obtained by computing the mean donation rate for observations that fall within the corresponding bin.

donation even controlling for time invariant attributes about the individual. Our field results were consistent with a theory that changes in beliefs about being pivotal drive changes in donation rates; our lab results provide evidence that this is indeed part of the mechanism at work.

As a robustness check we can show that *pivotal* beliefs (p_{T-1}) matter above and beyond the total likelihood of reaching the threshold without own donation ($\sum_{j=T}^9 p_j j$).⁴⁷ If we include the total likelihood of reaching the threshold without own donation ($\sum_{j=T}^9 p_j j$) as a control variable we find that the coefficient on pivotal beliefs is not only still positive and significant, but also it doubles (from 0.22 to 0.48 using the full sample, and from 0.43 to 0.78 using only those with variation; see columns 5 and 6 of Table 10 in the Appendix). As an additional robustness check, for those subjects with single peaked beliefs we can identify if that peak (or modal belief) is on $T - 1$ peer donors.⁴⁸ If we use a dummy variable taking the value one when the highest peak is on $T - 1$ other donors as our measure of beliefs about

within subjects experiment.

⁴⁷For example if the threshold were 5, then it might matter what the collective probability mass on 5, 6, 7, 8, and 9 other donors was, rather than the probability that only 4 people give. One way to think of this is that a person wants to donate if they think the threshold will be met without them, and they want to keep their money if they think the threshold won't be met without them. So that this person wants to be on the “winning” side regardless of whether that is threshold met or threshold not met.

⁴⁸For example if the threshold were 5, then we ask “Is the largest probability mass on p_4 ?”

being pivotal, we find that it has a positive and significant relationship with the likelihood of giving (see columns 7 and 8 of Table 10 in the Appendix).

In the field we found that different threshold match treatments (T1/T2/T3) had different effects on the donation rate with T3 doubling the donation rate from 1.6% to 3.6%. Our theoretical model implies that the effectiveness of T3 is attributable to changes in the beliefs about being pivotal in this treatment. Now using data from our lab study we have shown once again that T3 is highly effective in increasing the donation rate, but moreover that pivotal beliefs do indeed influence giving behavior. Higher beliefs about being pivotal, in any treatment, are associated with a higher likelihood of donation.

5 Alternative Explanations

In this section, we show that the literature’s existing explanations for individual behavior under a match have no explanatory power for our results.⁴⁹ However, an explanation involving beliefs about peers does explain what we observe in the field and the laboratory.

Table 5: Predictions of Various Explanations Regarding Treatment Effect of T3

#	Explanation	Treatment Effect of T3 (relative to Control)	Comments
1	Price effect	0	price of giving unaffected
2	Public good level effect	-	(potentially) increase in G
3	Signaling	0	excluded by design
4	Fictitious donor	0, -	
5	Donor will give anyway	0, -	
6	Warm glow	0	
7	Beliefs about peers	+	(with certain beliefs)

+ indicates a prediction of increase in response rate.

- indicates a prediction of decrease in response rate.

0 indicates a prediction of no change in response rate.

Karlan and List (2007) summarize six explanations that have been proposed for how matching might (or might not) affect donation decisions. The first is related to price effects because a dollar-for-dollar match can be interpreted as lowering the price of giving. Second, matching may reduce the marginal utility of contributing by increasing the level of the public

⁴⁹These existing explanations have often been considered in regards to a linear match setting, so it may not be surprising that they are unable to explain effects in our threshold match setting.

good that is provided, thereby resulting in decreased donations. Third is Vesterlund’s (2003) signaling explanation: a matching commitment may provide information to individuals that the charity is of high quality.⁵⁰ Fourth is the idea that individuals may discount the matching opportunity as fictitious, which could perhaps even harm the charity’s efforts via a negative reputation effect. The fifth explanation is that individuals believe that the matching third party donor exists but will give the full matching gift regardless of what other donations are received. The sixth involves the impure altruism model of Andreoni (1989, 1990), extended to large pools of individuals or a large size of the charity (Ribar and Wilhelm, 2002; Andreoni, 2006). In these cases, giving behavior is determined only by “warm glow” utility from giving, so contribution behavior is insensitive to the matching grant.

We consider these explanations in the context of our experimental results in the field and laboratory; Table 5 provides a summary. Do any of these explanations predict the observed large, positive, and statistically significant treatment effect of T3 relative to control? The traditional price effect story does not because the T3 treatment does not change the price of giving.⁵¹ The second explanation - that matching money reduces the marginal utility of giving by increasing the level of the public good - inaccurately predicts a reduction in giving.

Third, our experimental design rules out signaling as an explanation. Quality signaling cannot explain the T3 treatment effect in our field experiment, because solicited individuals were alumni who had experienced the quality of the educational program personally. Moreover, the matching amount of \$50 is unlikely to be interpreted as a strong signal of quality. An alternative story is urgency signaling, i.e., the match could indicate that “now” is a particularly important time to give (Karlan and List, 2007). As above, \$50 is unlikely to be interpreted as a strong signal of urgency. In addition, there is no supporting language in the field study letter to indicate that donations in the current window are any more important than in previous donation appeals. Moreover, in the lab experiment, all subjects face a choice of giving to the same charity both with and without a match, so there is no signaling value in the match about the charity’s quality or urgency of need. Fourth is the possibility that individuals perceive the matching opportunity as fictitious, and fifth is the explanation that individuals believe that matching funds will be given regardless of their actions. Neither of these can accommodate the increased donation rate in the T3 treatment.

⁵⁰Interestingly, Rondeau and List (2008) find in their Sierra Club experiment that an unconditional “challenge” gift is more effective at raising funds than a conditional matching commitment. Still, it is plausible that even conditional matching gifts provide a signal to individuals.

⁵¹A price effect changes the rate at which each dollar contributed translates into dollars received by the charity. That is, it is a marginal effect that applies to each dollar donated; thus, it operates along the intensive margin. In contrast, the match in treatment T3 involves a discontinuous effect that applies to the first dollar donated, but only if a sufficient number of others have donated. Thus, it operates along the extensive margin.

Sixth, the theory of warm glow giving also fails to explain the difference, given its prediction of individual insensitivity to matching funds.

We do not interpret our results as a rejection of these explanations in general.⁵² Rather, we interpret our results as identifying the importance of beliefs about peers, and we suggest that this factor may have explanatory power for donation behavior under more general forms of donation matching. For example, with a dollar-for-dollar match, an individual may believe that peers will exhaust the match, yielding less incentive to contribute relative to the scenario in which the individual believes her donation will secure matching money.⁵³

One potential concern is whether the treatment effects observed in the field experiment generalize to other populations. We provide a check on this concern by observing the treatment effect of the dollar-for-dollar match (M). This treatment is similar to those used in previously executed designs in the literature on a wide variety of populations. Treatment M's 1:1 dollar-for-dollar matching rate is consistent with the designs of Karlan and List (2007) and Karlan et al. (2011), while the \$3,000 of matching money offered to 600 people is similar to, but slightly more generous than, the \$2,500 matching dollars offered to 750 people in Rondeau and List (2008).⁵⁴ Our dollar-for-dollar match had a donation rate of 2.34%, which is quite close to those found by previous similar designs (Karlan and List, 2007; Rondeau and List, 2008). Thus, this check provides no evidence to suggest that this population varies from other populations in ways that would affect generalizability.

6 Conclusion

Donation matching is popular, yet field experiments find that matching does not always increase donations. This may occur because individuals believe that peers will exhaust the matching funds. This paper uses theory coupled with field and laboratory experiments to show that individuals are more likely to donate when they believe they are pivotal to securing matching funds. We show this first through a theoretical model that predicts that higher beliefs of being pivotal to securing matching funds are associated with a higher likelihood of donation. Next, we test this prediction using experiments in both the field and laboratory. Taken together, our experimental results show that: (1) higher beliefs of

⁵²For example, signaling remains a highly plausible explanation for the efficacy of large dollar-for-dollar matching funds, regardless of its inability to explain our results.

⁵³Multiple explanations may play a role here. Even if an individual believes that peers will exhaust the match, the individual may still see the match as a signal of quality or urgency.

⁵⁴However, the Rondeau and List (2008) design included a money-back refund if a specified threshold was not achieved. Other donation matches from the field experiment literature do not inform the individual of the number of other people who qualify for the match. Thus, individuals are left to form their own beliefs about this variable.

being pivotal to securing matching funds are indeed associated with higher donation rates, and (2) improvements in donation rates can be realized by structuring donation matches in innovative ways.

The results of our experiments hold interesting implications for fundraising practitioners. The most effective treatment in increasing donation rates - in both the field and laboratory - is a novel fundraising approach that offers a flat \$50 of matching money if at least 3 out of 10 individuals choose to donate. This treatment more than doubles the donation rate relative to control in both the field and laboratory.⁵⁵ This 3 out of 10 treatment also yields a donation rate that is substantially greater than a traditional dollar-for-dollar match in the field, though estimate imprecision prevents a finding of statistical significance. We interpret these results as support for the notion that dollar-for-dollar matching may not be the best use of matching funds.

At this point, these results can say little about effects of the studied treatments on long-run giving outcomes. The decision to give charitably may have an important dynamic component (Landry, Lange, List, Price, and Rupp, 2010). For example, individuals may set annual goals for their giving. In this case, increased giving during one appeal window may be accounted for by lowering giving in a later appeal. Along these lines, Meier (2007) presents a finding that matching may undermine the willingness to give charitably in the long run. Additionally, our results cannot speak to the ideal amount of information disclosure during a matching campaign. This experiment did not test how disclosing information about the total size of the donor pool, or other's previous or current donations (Shang and Croson, 2009; Croson and Shang, 2013), would affect the contributions. We have shown that beliefs about peers matter. Intuitively these beliefs could be influenced by the information revealed, so this is an important area for future research.

We have shown that beliefs about peers are likely to be an important determinant in charitable donation decisions when matching funds are available. We have also shown that in both our field experiment and lab experiment a "3 out of 10" threshold match doubles the donation rate when compared to the control. This threshold match procedure is more effective than a traditional dollar-for-dollar matching procedure in the field. To better understand the mechanism at work, we use a lab study that replicates the field results and further tests the importance of beliefs about peers. Beliefs about peers explain the effectiveness of

⁵⁵Our experiment uses groups of 10 people and a bonus of \$50 to find that T3 doubles the donation rate versus the control. This doubling may not hold if we changed either group size or bonus size, but believe this is an excellent path for future research. However, the exciting thing is that the doubling effect of the T3 treatment versus control holds for both the field with a very low base rate giving (1.6%) as well as the lab with the higher base rate of 24%. Larger groups have a more difficult time providing threshold public goods (Bagnoli and McKee, 1991), so it is likely the magnitude of our results might decrease with groups larger than 10.

our threshold match in both the field and laboratory. Further study of beliefs about peers may be insightful for both academics and practitioners.

References

- Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy* 97(6), 1447–1458.
- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal* 100(401), 464–477.
- Andreoni, J. (2006). *Philanthropy*, Volume 1 of *Handbook on the Economics of Giving, Reciprocity and Altruism*, Chapter 18, pp. 1201–1269. Elsevier.
- Anik, L., M. I. Norton, and D. Ariely (2014). Contingent match incentives increase donations. *Journal of Marketing Research*.
- Bagnoli, M. and M. McKee (1991). Voluntary contribution games: Efficient private provision of public goods. *Economic Inquiry* 29(2), 351–366.
- Blanco, M., D. Engelmann, A. Koch, and H.-T. Normann (2010). Belief elicitation in experiments: Is there a hedging problem? *Experimental Economics* 13(4), 412–438.
- Bruhn, M. and D. McKenzie (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 200–232.
- Charness, G., U. Gneezy, and M. A. Kuhn (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization* 81(1), 1–8.
- Charness, G. and P. Holder (2016). Charity in the laboratory: Matching, competition, and group identity. *Working Paper*.
- Chen, X.-P., W. T. Au, and S. Komorita (1996). Sequential choice in a step-level public goods dilemma: The effects of criticality and uncertainty. *Organizational Behavior and Human Decision Processes* 65(1), 37–47.
- Chen, Y., X. Li, and J. K. MacKie-Mason (2006). Online fund-raising mechanisms: A field experiment. *Contributions in Economic Analysis & Policy* 5(2).
- Crawford, V. P. and N. Iriberry (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions? *Econometrica* 75(6), 1721–1770.

- Croson, R. (2007). Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry* 45(2), 199–216.
- Croson, R. and J. Shang (2013). Limits of the effect of social information on the voluntary provision of public goods: evidence from field experiments. *Economic Inquiry* 51(1), 473–477.
- Diekmann, A. (1985). Volunteer’s dilemma. *Journal of Conflict Resolution*, 605–610.
- Eckel, C. and P. Grossman (2008). Subsidizing charitable contributions: a natural field experiment comparing matching and rebate subsidies. *Experimental Economics* 11(3), 234–252.
- Eil, D. and J. M. Rao (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics* 3(2), pp. 114–138.
- Exley, C. L. (2015). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies*.
- Fehr, E. and A. Leibbrandt (2011). A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics* 95(9), 1144–1155.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Fischbacher, U. and S. Gächter (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review* 100(1), 541–556.
- GivingUSA (2013). Giving usa 2014 highlights.
- Gneezy, U., E. A. Keenan, and A. Gneezy (2014). Avoiding overhead aversion in charity. *Science* 346(6209), 632–635.
- Goeree, J. K. and C. A. Holt (2004). A model of noisy introspection. *Games and Economic Behavior* 46(2), 365–382.
- Goeree, J. K. and C. A. Holt (2005). An explanation of anomalous behavior in models of political participation. *American Political Science Review* 99(2), 201–213.
- Holt, C. and A. Smith (2013). Belief elicitation with a simple lottery choice menu. *Working paper*.

- Huck, S. and I. Rasul (2011). Matched fundraising: Evidence from a natural field experiment. *Journal of Public Economics* 95(56), 351 – 362. Charitable Giving and Fundraising Special Issue.
- 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(2), 326–69.
- Isaac, M., D. Schmidt, and J. Walker (1989). The assurance problem in a laboratory market. *Public Choice* 62(3), 217–236.
- Karlan, D. and J. A. List (2007). Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *The American Economic Review*, 1774–1793.
- Karlan, D., J. A. List, and E. Shafir (2011). Small matches and charitable giving: Evidence from a natural field experiment. *Journal of Public Economics* 95(5), 344–350.
- Kessler, J. B. (2013). Announcements of support and public good provision. *Working Paper*.
- Landry, C. E., A. Lange, J. A. List, M. K. Price, and N. G. Rupp (2006). Toward an understanding of the economics of charity: Evidence from a field experiment. *The Quarterly Journal of Economics* 121(2), 747–782.
- Landry, C. E., A. Lange, J. A. List, M. K. Price, and N. G. Rupp (2010). Is a donor in hand better than two in the bush? Evidence from a natural field experiment. *The American Economic Review* 100(3).
- Manski, C. F. and C. Neri (2013). First-and second-order subjective expectations in strategic decision-making: Experimental evidence. *Games and Economic Behavior* 81, 232–254.
- McBride, M. (2010). Threshold uncertainty in discrete public good games: an experimental study. *Economics of governance* 11(1), 77–99.
- Meer, J. (2013). The habit of giving. *Economic Inquiry* 51(4), 2002–2017.
- Meier, S. (2007). Do subsidies increase charitable giving in the long run? Matching donations in a field experiment. *Journal of the European Economic Association* 5(6), 1203–1222.
- Moore, D. A. and P. J. Healy (2008). The trouble with overconfidence. *Psychological review* 115(2), 502.
- Myatt, D. P. and C. Wallace (2008). When does one bad apple spoil the barrel? an evolutionary analysis of collective action. *The Review of Economic Studies* 75(2), 499–527.

- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review* 85(5), 1313–1326.
- Rapoport, A. and D. Eshed-Levy (1989). Provision of step-level public goods: Effects of greed and fear of being gypped. *Organizational Behavior and Human Decision Processes* 44(3), 325–344.
- Ribar, D. C. and M. O. Wilhelm (2002). Altruistic and joy-of-giving motivations in charitable behavior. *Journal of Political Economy* 110(2), 425–457.
- Rondeau, D. and J. A. List (2008). Matching and challenge gifts to charity: Evidence from laboratory and natural field experiments. *Experimental Economics* 11(3), 253–267.
- Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics* 1(1), 43–62.
- Shang, J. and R. Croson (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal* 119(540), 1422–1439.
- Smith, S., F. Windmeijer, and E. Wright (2014). Peer effects in charitable giving: Evidence from the (running) field. *The Economic Journal*.
- Stahl, D. O. and P. W. Wilson (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization* 25(3), 309–327.
- Van de Kragt, A. J., J. M. Orbell, and R. M. Dawes (1983). The minimal contributing set as a solution to public goods problems. *American Political Science Review* 77(01), 112–122.
- Vesterlund, L. (2003). The informational value of sequential fundraising. *Journal of Public Economics* 87(3), 627–657.

Appendix (for Online Publication)

Table 6: Field Experiment Balance on Pre-Experiment Attributes

Treatment	Cohort Year	Warm4	Male	Count Prev. Donations	Sum Prev. Donations	Mean Prev. Donations	Days Since Last Donation
C	1993.41	0.11	0.51	0.73	130.92	38.45	434.66
M	1992.32	0.10	0.48	0.97	105.50	28.94	521.78
T1	1992.30	0.11	0.48	0.74	89.20	23.76	518.85
T2	1992.31	0.11	0.47	0.75	84.32	22.84	460.03
T3	1992.30	0.11	0.47	0.97	80.85	16.19	391.58
F Statistic	2.35*	0.11	0.73	1.38	0.33	0.31	0.26
Min	1977	0	0	0	0	0	0
Max	2007	1	1	53	35,000	17,500	7,112

Table 7: Field Experiment Dependent Variable Probability of Donation (0 to 100)

	In Text Warm4 1	Logit 2	Warm1 3	Warm2 4	Warm3 5	Warm5 6
Match (M)	0.731 (0.829)	0.446 (0.486)	0.965 (0.832)	1.117 (0.855)	0.822 (0.830)	0.591 (0.830)
Threshold Match 1 (T1)	0.335 (0.975)	0.289 (0.608)	0.578 (1.019)	0.686 (0.987)	0.532 (0.967)	0.417 (0.989)
Threshold Match 2 (T2)	0.675 (0.923)	0.491 (0.510)	0.878 (0.897)	1.129 (0.877)	0.856 (0.916)	0.833 (0.937)
Threshold Match 3 (T3)	1.945* (1.051)	1.017** (0.488)	2.147* (1.066)	2.116* (1.057)	1.999* (1.041)	2.153* (1.095)
warm1			25.752*** (4.408)			
warm2				21.657*** (4.002)		
warm3					17.607*** (3.104)	
warm4	15.876*** (2.694)	3.504*** (0.442)				
warm5						13.286*** (2.279)
Adjusted R2	0.107		0.121	0.129	0.114	0.089
Pseudo R2		0.305				
N	2,567	2,120	2,567	2,567	2,567	2,567

Notes: All coefficients are multiplied by 100 for ease of readability. Standard errors are clustered at the cohort level. Observations are weighted to account for the constraints on randomization. Legend: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table 8: Lab Experiment: Realized vs. Expected Number of Peer Donors

Treatment	Realized Peer Donors	Expected Peer Donors
Control	2.16	2.11
T1	3.51	3.56
T3	4.50	3.82
T5	3.87	3.91
T7	4.05	3.90
T10	3.78	3.95

Notes: There were 100 subjects who each participated in 6 treatments. The realized and expected number of peer donors can take a value between 0 other donors up to 9 other donors, so it does not include the subject's own choice to keep or donate.

Table 9: Lab Experiment Logit Models Dependent Variable Probability of Donation (0 to 100)

	With Control 1	With Beliefs 2	Without Beliefs 3
T1	1.33*** (0.46)		
T3	2.11*** (0.47)	0.84** (0.36)	0.77** (0.35)
T5	1.62*** (0.46)	0.28 (0.41)	0.29 (0.38)
T7	1.76*** (0.45)	0.51 (0.48)	0.43 (0.46)
T10	1.55*** (0.39)	0.09 (0.45)	0.22 (0.43)
Pivotal Belief		0.02*** (0.01)	
Pseudo R2	0.10	0.05	0.02
N	366	285	285

Notes: All coefficients are multiplied by 100 for ease of readability. All models include subject fixed effects and standard errors are clustered at the subject level.
Legend: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table 10: Alternative Specifications for Lab Experiment Dependent Variable: Probability of Donation (0 to 100)

	<u>In Text</u>		<u>Without Beliefs</u>		<u>With Total Prob.</u>		<u>Highest Peak Beliefs</u>	
	<u>Full</u> <u>1</u>	<u>Variation</u> <u>2</u>	<u>Full</u> <u>3</u>	<u>Variation</u> <u>4</u>	<u>Full</u> <u>5</u>	<u>Variation</u> <u>6</u>	<u>Full</u> <u>7</u>	<u>Variation</u> <u>8</u>
Pivotal Belief Level	0.22** (0.09)	0.43*** (0.15)			0.48*** (0.10)	0.78*** (0.15)		
Total Prob. Threshold Met					0.66*** (0.10)	1.07*** (0.13)		
$T - 1$ Is Highest Peak							14.17** (6.27)	23.35** (9.68)
T3	10.71* (5.56)	20.78** (9.59)	11.00** (5.50)	19.30** (9.53)	21.66*** (5.74)	32.73*** (9.26)	14.66** (5.98)	26.65** (10.48)
T5	3.08 (6.17)	6.22 (11.07)	4.00 (5.95)	7.02 (10.49)	32.27*** (7.11)	48.99*** (9.28)	4.62 (6.95)	6.72 (13.20)
T7	6.58 (7.21)	12.01 (12.81)	6.00 (7.10)	10.53 (12.50)	51.09*** (9.83)	80.09*** (12.82)	8.18 (7.69)	13.68 (14.41)
T10	1.71 (6.67)	2.80 (11.75)	3.00 (6.67)	5.26 (11.76)	57.72*** (10.39)	89.61*** (13.04)	2.14 (7.35)	2.09 (13.82)
constant	36.01*** (4.35)	29.76*** (7.67)	39.00*** (4.20)	36.84*** (7.37)	-24.76** (10.44)	-65.05*** (12.22)	37.21*** (4.82)	31.95*** (8.42)
R2	0.54	0.22	0.54	0.20	0.60	0.38	0.62	0.33
Adj R2	0.42	0.01	0.42	-0.01	0.50	0.21	0.50	0.08
N	500	285	500	285	500	285	413	223
Subjects	100	57	100	57	100	57	99	56

Notes: All coefficients are multiplied by 100 for ease of readability. All models include subject fixed effects, and standard errors are clustered at the subject level. In columns 7 and 8 we restrict the sample to observations where a subject reported single peaked beliefs. Legend: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

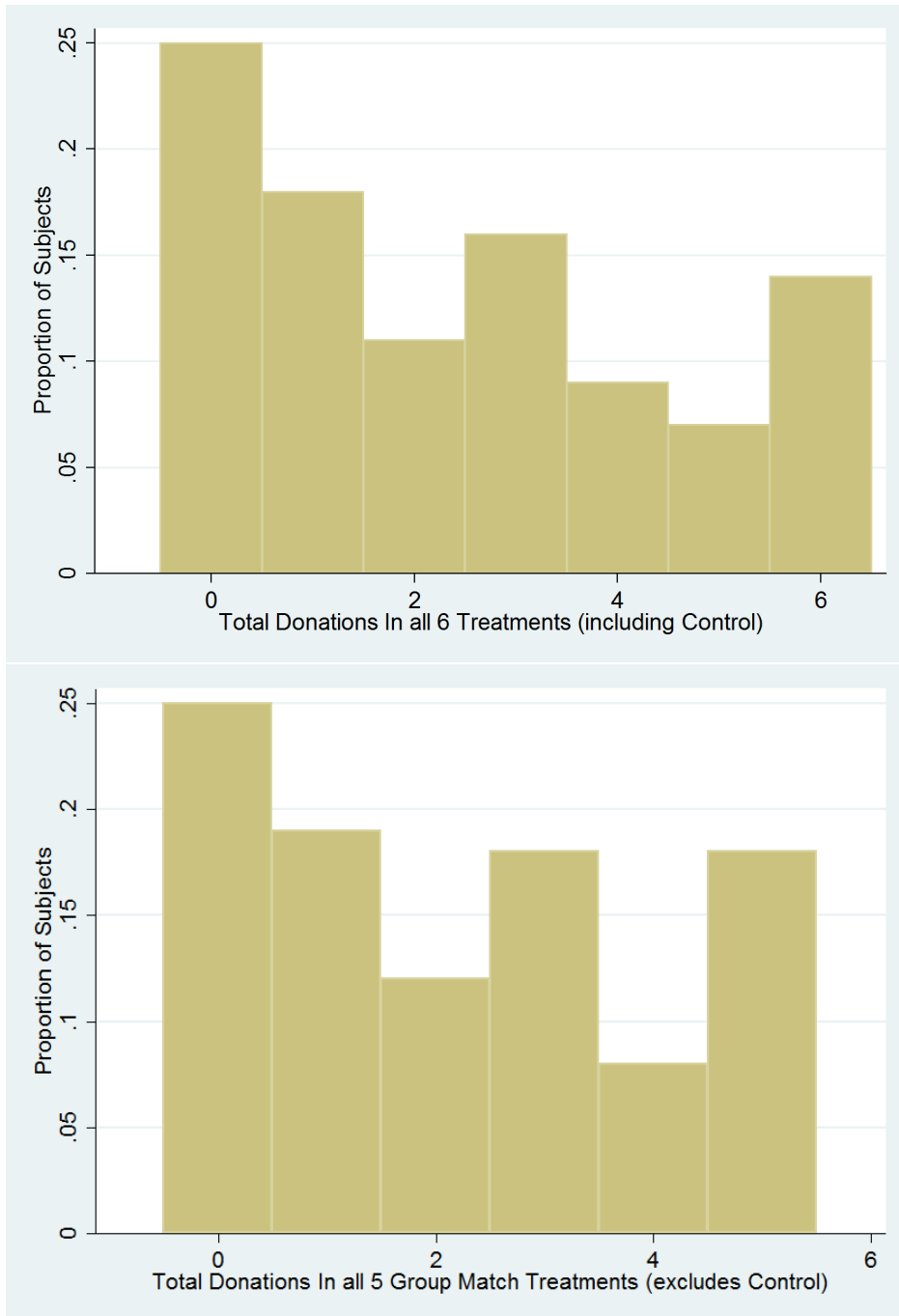


Figure 4: Histogram of Donation Counts per Subject

Note: This histogram indicates the number of subjects associated with a certain number of total donations in the experiment. The top panel shows data for all 6 treatments, while the bottom panel shows data for only the Threshold Match treatments.

6.1 Comparative Static on Belief of Being Pivotal, Regardless of Risk Type

This proof demonstrates that the comparative static on p_{T-1} holds for all risk types.

There are two decision options: contribute (C) or keep (K). For each of those options, there are two possible outcomes: the threshold is reached, or it is not. Thus, there are ultimately 4 payoff values that can be realized:

- Contribute and the threshold is achieved (call this \bar{C})
- Contribute and the threshold is not achieved (call this \underline{C})
- Keep and the threshold is achieved (call this \bar{K})
- Keep and the threshold is not achieved (call this \underline{K})

We have $\bar{K} > \underline{K}$ and $\bar{C} > \underline{C}$ because conditional on choice, if the threshold is achieved then the agent receives the associated positive payoff. If $\underline{K} > \bar{C}$, then keeping is a dominant strategy regardless of risk type. The remainder of the proof focuses on the condition $\bar{C} > \underline{K}$.

Define Bernoulli utility function $u(\cdot)$ over payoffs. The function is increasing.

The expected utility of contributing is given by:

$$EU_{contribute} = \sum_{j=0}^{T-2} p_j u(\underline{C}) + p_{T-1} u(\bar{C}) + \sum_{j=T}^{N-1} p_j u(\bar{C})$$

where p_j denotes the probability that j of the $N - 1$ peer donors choose to donate. The expected utility of keeping is given by:

$$EU_{keep} = \sum_{j=0}^{T-2} p_j u(\underline{K}) + p_{T-1} u(\underline{K}) + \sum_{j=T}^{N-1} p_j u(\bar{K})$$

It is optimal to donate if the expected utility of contributing exceeds the expected utility of keeping, which occurs if:

$$p_{T-1} [u(\bar{C}) - u(\underline{K})] > \sum_{j=0}^{T-2} p_j [u(\underline{K}) - u(\underline{C})] + \sum_{j=T}^{N-1} p_j [u(\bar{K}) - u(\bar{C})]$$

This has a marginal benefit / marginal cost interpretation. The LHS is the marginal benefit of contributing: with p_{T-1} the donation is pivotal and the individual realizes payoff

\bar{C} instead of \underline{K} . The RHS is the marginal cost of contributing: when the donation is not pivotal, the individual sacrifices private consumption but gets the warm glow.

We now consider a comparative static on p_{T-1} . An increase in this value must be accompanied by a decrease in at least one other p_j , because $\sum_{j=0}^{N-1} p_j = 1$. So an increase in p_{T-1} not only increases the LHS (recall $\bar{C} > \underline{K}$), but it reduces the RHS when $[u(\underline{K}) - u(\underline{C})]$ and $[u(\bar{K}) - u(\bar{C})]$ are both positive. If those quantities are both negative, then the dominant strategy is to contribute regardless of the value of p_{T-1} because the RHS is negative (this is the scenario where the warm glow of giving is so large it outweighs the cost of giving up private consumption by donating).

Therefore we have shown that as long as $u(\cdot)$ is increasing, an increase in p_{T-1} makes the individual more likely to donate (unless the individual's dominant strategy is to contribute, in which case contributing remains the dominant strategy). This holds for risk neutral, risk averse, and risk loving preferences.