



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

No Substitute for the Real Thing: The Importance of In-Context Field Experiments in Fundraising

Indranil Goswami Oleg Urminsky

To cite this article:

Indranil Goswami Oleg Urminsky (2020) No Substitute for the Real Thing: The Importance of In-Context Field Experiments in Fundraising. Marketing Science 39(6):1052-1070. <https://doi.org/10.1287/mksc.2020.1252>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2020, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

No Substitute for the Real Thing: The Importance of In-Context Field Experiments in Fundraising

Indranil Goswami,^a Oleg Urminsky^b

^a School of Management, State University of New York at Buffalo, Buffalo, New York 14260; ^b Booth School of Business, University of Chicago, Chicago, Illinois 60637

Contact: goswami4@buffalo.edu,  <https://orcid.org/0000-0002-7477-5525> (IG); oleg.urminsky@chicagobooth.edu,  <https://orcid.org/0000-0003-4390-386X> (OU)

Received: October 2, 2017

Revised: September 21, 2018; January 5, 2020

Accepted: May 12, 2020

Published Online in Articles in Advance:
October 6, 2020

<https://doi.org/10.1287/mksc.2020.1252>

Copyright: © 2020 INFORMS

Abstract. We present a complete empirical case study of fundraising campaign decisions that demonstrates the unique importance of in-context field experiments. We first design novel matching-based fundraising appeals. We then discuss the assumptions needed to derive theory-based predictions from the standard impure altruism model and solicit expert opinion about the potential performance of our interventions. Both theory-based and experts' predictions suggested improved fundraising performance from framing a matching intervention as crediting donors for the matched funds, whereas predictions for the other appeals were more ambiguous. However, the results of a natural field experiment with prior donors of a nonprofit instead showed a significantly poorer performance from employing the giving-credit framing. This surprising finding was replicated in a second natural field experiment to confirm the ground truth, at least within a specific context. Experts also lacked consensus about a conditional matching scheme, which, in fact, did not improve fundraising. Theoretically, our results highlight the limitations of both impure altruism models and expert opinion in predicting complex warm-glow motivation. More practically, our results question the availability of useful guidance and suggest the indispensability of field testing for behavioral interventions in fundraising.

History: This paper has been accepted for the *Marketing Science* Special Issue on Field Experiments.

Supplemental Material: Data and the online appendix are available at <https://doi.org/10.1287/mksc.2020.1252>.

Keywords: donation • matching • altruism • warm glow • motivation • field experiment

1. Introduction

One of the most common solicitation techniques used in fundraising is *matching*—communicating to prospective donors that an external donor has committed to making a contribution based on the amount given by prospective donors. A matching solicitation can be thought of as a conditional leadership gift that uses a commitment to match others' contributions at a given rate, sometimes limited to the maximum amount the leader is willing to give (Karlan and List 2007). *Standard matching* solicitations promise to match every dollar donated, typically using an equal-match ratio (i.e., a one-to-one ratio). Both higher match ratios (e.g., two-to-one, three-to-one, etc.; Dove 2000, Karlan and List 2007) and lower match ratios (e.g., one-to-three; Karlan et al. 2011) have also been used. Popular press reports describe matching as a “staple of fund-raising” (Leonhardt 2008), and a large body of academic work has studied matching solicitations (Andreoni and Payne 2013).

The empirical results from research on matching solicitations, however, have been extremely mixed.

There is some evidence that matching improves fundraising outcomes (Eckel and Grossman 2003; Karlan and List 2007, 2012; Meier 2007; Martin and Randal 2009; Bekkers 2015). At the same time, other papers have failed to find a positive effect or have even reported a negative effect of matching (Davis et al. 2005, Rondeau and List 2008, Baker et al. 2009, Huck and Rasul 2011, Karlan et al. 2011).

One particular concern about the effects of matching on funds raised is the potential for crowding-out behavior. Donors may scale back their contributions in response to a match offer that makes each dollar donated go farther (Huck et al. 2013, Adena and Huck 2017). Similarly, a higher match ratio (compared with a standard one-to-one match) does not always increase fundraising performance (Karlan and List 2007). Speculations in the literature attribute such behavior to scope insensitivity for quality signals (Karlan and List 2007) and lack of attention to price information (Eckel and Grossman 2017).

How can a fundraiser navigate the possibilities and decide on the best matching solicitation to use? What

sources of guidance should the fundraiser rely on to make an informed decision about the offer design? In this paper, we show that commonly available sources of guidance, both the implications of theoretical models and the opinions of experts, can fall short of correctly predicting the results of novel fundraising tactics. As a result, in-context field experiments appear to be indispensable for identifying optimal strategies in fundraising.

We explore two types of novel matching interventions: reframing who gets credit for matched donation funds and setting a cutoff or a threshold below which donated funds are not matched. We first turn to a basic model of *warm glow* and find that the model is underspecified to derive direct predictions as to the efficacy of interventions. However, we argue that under reasonable additional assumptions, the model suggests that appeals using the novel framing intervention may improve fundraising. Next, we present the advice of experienced fund managers, collected in two surveys, which largely echoes both the predictions and basis for reasoning of the model. Finally, we present tests of whether these predictions match actual donor behavior in two separate natural field studies in which decision makers were unaware that they were participating in an experiment. The findings from the field reveal markedly opposite results: the novel framing interventions that were deemed most promising raised significantly less money than the regular framing.

The goal of this paper is to assess the adequacy of available sources of policy guidance in contrast with in-context field experiments as the source of *ground truth*, focusing on the domain of fundraising. Given the myriad and complex ways in which framing, cognitive processes, and contextual influences can affect warm-glow preferences, our results highlight the limited usefulness of model derivations and expert opinions in correctly anticipating the actual effects of fundraising interventions. We demonstrate that the relevant theoretical model of impure altruism (Andreoni 1990) represents a useful framework that can accommodate many alternative outcomes but does not generally make unambiguous predictions needed for decision making without making significant additional assumptions.

If experts reason according to the theoretical models but also have the experience to choose reasonable assumptions to fill in the gaps in the theoretical model, experts may make the accurate predictions needed to correctly choose optimal interventions. However, our results suggest that although experts' reasoning does correspond to the key elements of the theoretical model, experts were unable to accurately predict the outcomes. Therefore, we conclude that although both

model predictions and expert advice can provide a useful initial starting point for designing charity appeals, these approaches are no replacement for in-context field experiments to determine the outcomes of behavioral interventions.

This paper proceeds as follows. In Section 2, we introduce our proposed interventions and assess assumptions under which current theoretical models of impure altruism make predictions for the outcomes of these interventions. In Sections 3 and 4, we report the predictions of professional fundraising managers and highlight general convergence of guidance obtained from these two sources. Next, we contrast these predictions with the actual results of two natural field experiments, the second a preregistered confirmatory replication of the first (Sections 5 and 6). Our results represent a self-contained empirical case study demonstrating that the existing academic guidance and fundraisers' own experience-based intuitions poorly anticipate the real outcomes in the field. We conclude in Section 7 with a discussion of the implications of our findings for theories of altruism, for the validity of expert opinion, and for fundraising practices.

2. Predictions from Theoretical Models as Initial Guidance to the Fundraiser

The standard model of altruistic behavior in fundraising, based on Andreoni (1990; also see Karlan and List 2007), can be written as

$$U_i = F(y_i, G, g_i). \quad (1)$$

Here an individual agent i gets utility from private consumption y_i , the total quantity of funds G raised by a charity, and the agent's own private contribution to charity g_i . Assuming that y_i , G , and g_i are all normalized to the same units, for example, dollars, the consumption bundle (y_i, g_i) is related by the budget constraint $y_i + g_i \leq w_i$. Assuming that utilities are additively separable, the total utility can be written as

$$U_i = u(y_i) + \delta_i h(G + g_i) + \gamma_i f(g_i). \quad (2)$$

It is typically assumed that $u(\cdot)$, $h(\cdot)$, and $f(\cdot)$ are identical across people and that each is concave and increasing in y_i , G , and g_i , respectively (Lange et al. 2007). The component $\delta_i h(G + g_i)$ represents pure altruism—the utility a person receives from the charity having the funds—which depends on the individual's weight on pure altruism δ_i (e.g., which might, in turn, depend on factors such as beliefs about the quality of the organization; List and Lucking-Reiley 2002). The component $\gamma_i f(g_i)$ represents the individual-specific utility, or *warm glow*, that a donor receives from personally contributing funds over and above the utility of the charity having those funds. The individual

experiences more warm glow if he or she personally donates a larger amount g_i or if other factors increase the importance γ_i of his or her donation (e.g., boosting the ego (Andreoni 1990) or emphasizing personal donor benefits (List et al. 2019)). In deciding how much to donate, the benefits from pure altruism and warm glow are traded off against the cost because of the donor's loss of utility from foregone private consumption $u(y_i)$.

2.1. Matching Solicitations

Next, we present an extension of the standard impure altruism model that explicitly incorporates matching donations (adapted from Karlan and List 2006). Consider a $k : 1$ match such that every dollar donated to charity is matched by $\$k$ from an external donor (i.e., the *match funder*). In this case, a donation of g_i yields $(k + 1)g_i$ of funds for a multiplier of $\Phi = k + 1$. If potential donor i believes that n other individual donors will each give an average of g , the total amount of money raised by the n other donors, including the match, is $G = n\Phi g$. The utility of potential donor i when making private contribution g_i can then be expressed as

$$U_i = u(w_i - g_i) + \delta_i h(G + \Phi g_i) + \gamma_i f(g_i). \quad (3)$$

The first-order condition (FOC) for optimal individual giving under this model occurs when the marginal utility of foregone consumption is equal to the marginal benefit of donating from pure altruism and warm glow up to the budget constraint w :

$$u'(w_i - g_i) = \Phi \delta_i h'(G + \Phi g_i) + \gamma_i f'(g_i). \quad (4)$$

2.2. Prior Empirical Findings and the Standard Matching Solicitation Model

The FOC in (4) provides an important theoretical framework for understanding donation decisions and makes some specific predictions. In particular, when the matching multiplier Φ is higher or people derive higher utility from pure altruism (δ_i) or derive more utility from warm glow (γ_i) based on their own donations, they will donate more, all else equal (Meier and Frey 2004, Crumpler and Grossman 2008). However, flexibility in the functional forms, the fact that the decision-weight parameters (δ_i and γ_i) are not observed, and the challenge of translating differences in real-world fundraising contexts into the model collectively mean that the model may fail to generate unambiguous predictions for interventions of interest. In fact, recent empirical findings have raised questions about the seemingly obvious prediction that higher matching multipliers ($\Phi > 2$) will result in more funds raised (Karlan and List 2007, but see Meier and Frey 2004). The problem is that in real-world

contexts, “all else equal” may be a very strong assumption, and even a seemingly simple intervention, such as a higher match, may have other effects.

For example, a higher matching multiplier Φ may serve as a credible quality signal (Karlan and List 2007) either directly or by signaling more cooperation from others (Eckel and Grossman 2003, Bekkers 2015; also see Frey and Meier 2004, Shang and Croson 2009), thereby bolstering the pure altruism benefit δ_i . In contrast, a higher matching multiplier Φ can also suggest to a potential donor that the total funds raised will increase, either directly because of the multiplier or indirectly by increasing the number of other donors. When this happens, the anticipated marginal utility from pure altruism (e.g., from increasing the total amount raised by the charity) will decrease, and the individual's donation decision will then be relatively more driven by warm-glow preferences (Ribar and Wilhelm 2002, Crumpler and Grossman 2008). As a result, seeing a higher matching level might induce a donor to give less because the benefits of matching funds crowd out the benefit of one's own contribution, particularly if warm-glow benefits are relatively low.

Other research has suggested that factors in the donation context may independently affect how much warm-glow utility $\gamma_i f(g_i)$ people derive from their own donations g_i . For example, the nature of solicitor–solicitee interaction (DellaVigna et al. 2012), the attractiveness of the solicitor (Landry et al. 2006), generous acknowledgment and recognition of one's contribution (Harbaugh 1998), and psychological benefits of associating with the charity (e.g., Escalas and Bettman 2003) have been theorized to increase warm glow (e.g., by boosting γ_i) and, if so, to increase the benefits of personally giving. However, further complicating the issue, such interventions could also signal higher charity quality, which might have an ambiguous effect on the pure altruism benefit, as discussed earlier (i.e., it could increase the pure altruism benefit via δ_i or could decrease the marginal benefit of one's own donation if the donor believes that total funds raised from others G will increase). Furthermore, such interventions could potentially even reduce warm glow if seen as cynical attempts to manipulate donors.

Therefore, the impure altruism model generally does not make unambiguous predictions about how matching solicitations will affect fundraising compared with a no-match control treatment. The model typically makes straightforward predictions only under the all-else-equal assumption that a change in the match appeal will affect only one parameter and will not impact other parts of the model. When an intervention has multiple potentially conflicting

impacts on motivation, as described earlier for matching, the impure altruism model is particularly uninformative as to the net effects. This is particularly problematic for the literature's ability to inform fundraisers, who often use matching as a strategy to raise money (Dove 2000). Next, we introduce two novel variations of matching incentives and explore whether they would be predicted to be effective in increasing funds raised, first from the perspective of the academic literature (as captured in the impure altruism model of charitable giving) and then from the perspective of professions fundraisers' beliefs.

2.3. Predictions for Novel Matching Campaign Strategies

2.3.1. Giving Credit to the Donor Framing. First, we propose a *giving-credit* framing, in which we suggest to the donor that the match is being added to the *donor's* contribution instead of being made as a separate donation by the *match funder*. Previous research has demonstrated framing effects on donation decisions. In particular, rebate framing (e.g., give \$10 to the charity and get back \$5 from a third party) has been shown to underperform a financially equivalent match framing (e.g., give \$5 to the charity and a third party will give the charity a \$5 match; Eckel and Grossman 2003, Davis et al. 2005, Blumenthal et al. 2012). This has been attributed to a differential misunderstanding of the consequences of matching subsidies (Davis et al. 2005) or to differential beliefs about others' donations (Bekkers 2015). For these interventions, what the model would predict a priori depends on assumptions about how the framing is interpreted.

Our giving-credit framing is intended to operate as a mental accounting intervention (Zhang and Sussman 2018), adding the match amount to the donor's perceived own contribution g_i so that the match increases not only the pure altruism utility but also the private warm-glow utility. Framing-based shifts in warm glow have been tested by Tonin and Vlassopoulos (2010), who found increased motivation from informing workers that the more effort they chose to exert, the more of a predetermined £15 charity donation would be credited to them. These results suggest that our proposed new giving-credit framing could successfully prompt donors to incorporate the match into their perceived warm-glow utility. Should that occur, the first-order conditions for an individual donor would instead be

$$u'(w_i - g_i) = \Phi \delta_i h'(G + \Phi g_i) + \Phi \gamma_i f'(\Phi g_i). \quad (5)$$

Given that the individual donation g_i is likely to be small (with respect to the budget constraint w_i), we can assume $f(\cdot)$ to be weakly monotonically increasing in that range. In this case, compared with standard

matching (see Equation (4)), the giving-credit frame would increase incremental utility from donating via the multiplier effect of matching on warm-glow preferences, given that

$$\Phi \gamma_i > \gamma_i \text{ for } \Phi > 1. \quad (6)$$

To derive a model prediction about the effect on donor behavior, we need to appeal to the all-else-equal assumption that interventions have only the intended primary effect without secondary effects on other parameters, as discussed earlier. In this case, we need to assume that the giving-credit frame increases the *amount* of funds yielding warm glow but does not impact other parameters in the model, such as the pure altruism value δ_i , the weight on warm-glow utility γ_i , or believed funds raised G . Under this all-else-equal assumption, the warm-glow model of altruistic giving unambiguously predicts more funds raised with the credit-framed match than with a regularly framed match that does not result in the match being included in the perceived personal donation. Consequently, the giving-credit framing serves as a test case to examine the validity of guidance offered by the theoretical model about how interventions affect donations under the strong all-else-equal assumption. From a practical point of view, the proposed framing is costless, unlike raising the match ratio, and therefore, the proposed framing intervention has not only theoretical but also practical significance.

2.3.2. Threshold Matching Mechanism. Next we discuss a second potential intervention, a personalized donation amount threshold for matching, which has been proposed previously to reduce the potential crowding-out effects of a matching offer (e.g., Sanders et al. 2013). This proposal is broadly consistent with prior research suggesting that implementing a match threshold for the number of other donors (Anik et al. 2014) or total funds raised (Baker et al. 2009) has a beneficial effect on funds raised.

Specifically, we propose that prior donors will have their new donations matched for every dollar they contribute over and above their previous contributions. Defining p_i as donor i 's most recent prior donation, the individual's matching multiplier Φ_i now depends on his or her prior and current donations:

$$\Phi_i = 2 - \min\left(1, \frac{p_i}{g_i}\right). \quad (7)$$

Clearly, this is a poorer match than a full one-to-one match, for which Φ would be equal to two for all donors i . We define \bar{G} as a donor's belief about how much will be raised from other donors, including the incremental matches induced by those other donors'

contributions. Under the threshold match offer, and again making the all-else-equal assumption (that other parameters will not be affected by the threshold match), the FOC will be the same as that for no match when $g_i \leq p_i$. However, there will be a discontinuity at $g_i = p_i$ such that the FOC will differ from that for the full match for $g_i > p_i$:

$$u'(w_i - g_i) = \begin{cases} \delta_i h'(\bar{G} + g_i) + \gamma_i f'(g_i), & \text{if } g_i \leq p_i, \\ 2\delta_i h'(\bar{G} + 2g_i - p_i) + \gamma_i f'(g_i), & \text{if } g_i > p_i. \end{cases} \quad (8)$$

Compare this with the FOC for a standard match (Equation (4)). Giving less than or the same amount as in the prior year is less attractive under the threshold match than under a standard match because the match does not apply, and pure altruism utility will be lower. What remains to be considered is whether the marginal utility of giving an amount $g_i > p_i$ is higher with the standard match or the threshold match, that is,

$$\Phi \delta_i h'(G + \Phi g_i) + \gamma_i f'(g_i) \quad \text{versus} \\ 2\delta_i h'(\bar{G} + 2g_i - p_i) + \gamma_i f'(g_i).$$

Consider the case where $\Phi \geq 2$. With the threshold match (compared with the standard match), the multiplier of h' will be at least as large, and h' will be evaluated at a lower value as long as $p_i > 0$. Under the assumption that $h'(\cdot)$ is decreasing (i.e., $h(\cdot)$ is concave), the marginal utility of donating $g_i > p_i$ is therefore higher with the threshold match. This suggests that people who would donate under either match will give more under the threshold match. However, those who would have given $g_i < p_i$ under the standard match might not find enough utility from donation to give when not matched under the threshold match.

Thus, the impure altruism model does not yield an unambiguous prediction, and additional assumptions are required. If we further assume stable revealed preferences of prior donors (i.e., because they found positive utility from giving p_i in the past without a match, they would do so again), then the impure altruism model predicts that the threshold match will increase funds raised. Incorporating the earlier discussion on the giving-credit framing, the combination of the threshold match and giving-credit framing should yield more donations than the threshold match alone.

The model-based predictions we have now derived by extending the impure altruism model are subject to the all-else-equal assumption, which could fail. For example, the threshold match could be seen as stingy, potentially signaling lower charity quality and reducing δ_i . The threshold match or giving-credit framing could be seen as manipulative, potentially reducing γ_i , which

would reduce the warm-glow benefits. Furthermore, a donor's beliefs about the positive or negative impact of the threshold match or the giving-credit framing on other donors could result in higher or lower expectations about total funds raised from others (i.e., G or \bar{G}) than with the standard match. As a result, the model predictions are ambiguous without making the additional assumptions discussed in this section. Next, we turn from the predictions of the model to predictions from experienced professionals.

3. Study 1: Expert Opinion as Guidance to the Fundraiser

A different (and arguably more prevalent) source of guidance for charities than theoretical economic models is fundraisers' experience and expertise, whether the charity's own staff or peers with corresponding roles at other organizations. Indeed, researchers recommend collecting expert beliefs before conducting experiments (DellaVigna and Pope 2017).

How would experts' beliefs compare with the model predictions? The beliefs of fundraising experts could reflect the model implications, particularly if the model represents a good description of reality and experts are well calibrated. However, experts' beliefs could also diverge from the model predictions either because experts are less accurate than relatively accurate statistical models (Dawes and Corrigan 1974, Dawes et al. 1989) or because experts have learned about donors' actual behavior that in fact systematically diverges from the implications of an inaccurate model. To test this, we surveyed experts with practical experience in raising money for nonprofits to measure their beliefs about the causal effects of the proposed interventions. In particular, we elicited separate beliefs about participation rates and donation size among those who participate, because these outcomes can be differentially impacted by contextual factors (e.g., suggested donation size; Goswami and Urminsky 2016a).

3.1. Design

We used the services of a professional online panel company to recruit fundraising managers of nonprofit organizations ($n = 105$) for a brief survey. Participants had an average of 10.2 years of experience in fundraising-related work, and 66% of participants reported having worked in fundraising campaigns that specifically used matching contributions.

Experts read about a direct-mail fundraising campaign for a nonprofit and about the intended target audience (see Table 1 for a summary; the actual stimuli used in all studies are provided in the online appendix). Experts then read about each of the five different appeal messages discussed earlier: control (no

Table 1. Appeals with actual descriptions that were shown to experts in Study 1

Number	Appeal	Solicitation text shown to experts
1	Control	"During our 75th Anniversary, we hope you will continue to join us in demonstrating your commitment to Chicago's art and artists by making a contribution today."
2	Standard matching	"In recognition of the Organization's success over its 75-year history, a supporter has offered a matching grant to encourage you to increase your donation and invest in our future." "This supporter will give \$1 for EVERY \$1 you contribute. So, for every dollar you give , we will receive two dollars in support of our programs—your dollar and a dollar from this supporter." "Let's not lose this match—please give today!"
3	Standard matching with giving-credit framing	"In recognition of the Organization's success over its 75-year history, a supporter has offered a matching grant to encourage you to increase your donation and invest in our future." "This supporter will add \$1 to your contribution for EVERY \$1 you give. So, for each dollar you give, we will receive two on your behalf in support of our programs." "Let's not lose this match—please give today!"
4	Threshold matching	"In recognition of the Organization's success over its 75-year history, a supporter has offered a matching grant to encourage you to increase your donation and invest in our future." "This supporter will give \$1 for every additional \$1 you donate OVER your last gift . So, for each dollar you add to the amount of your last contribution of '\$PREVIOUS,' we will receive two in support of our programs—your dollar and a dollar from this supporter." "Let's not lose this match—please give today!"
5	Threshold matching with giving-credit framing	"In recognition of the organization's success over its 75-year history, a supporter has offered a matching grant to encourage you to increase your donation and invest in our future." "This supporter will add \$1 to your contribution for every \$1 you donate OVER your last gift . So, for every dollar you add to the amount of your last contribution of '\$PREVIOUS,' we will receive two dollars on your behalf in support of our programs." "Let's not lose this match—please give today!"

Note. The placeholder "\$PREVIOUS" in appeals 4 and 5 reminded the donor about his or her last contribution amount.

matching), standard matching (i.e., one-to-one matching), standard matching with a giving-credit framing, threshold matching (i.e., one-to-one matching above the prior amount), and threshold matching with a giving-credit framing.

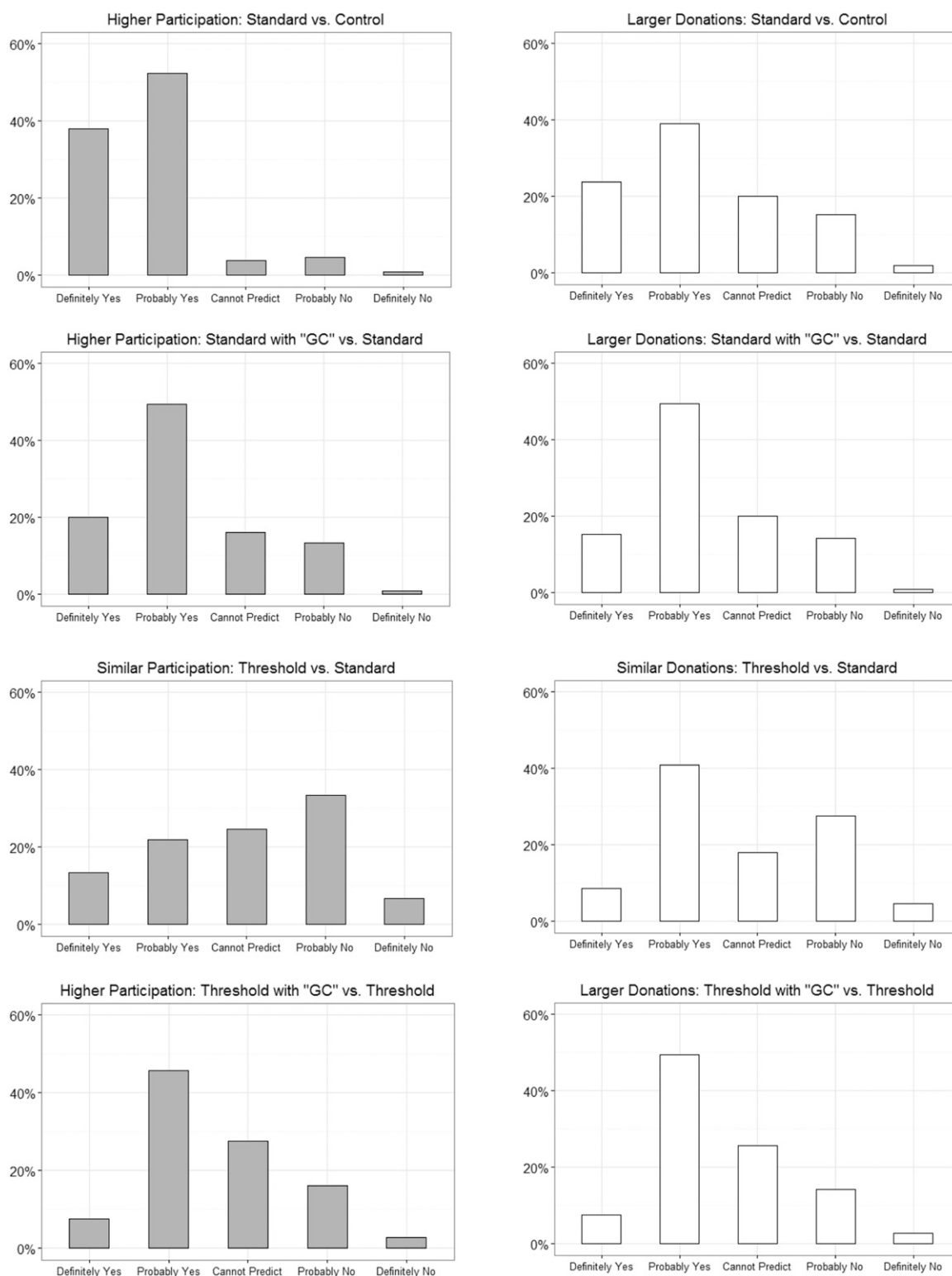
This within-subjects design elicits expert advice for multiple fundraising appeals simultaneously, which is similar to the approach used in recent research (e.g., Della Vigna and Pope 2017). Although this is different from how potential donors typically encounter appeals (one at a time), it realistically reflects the common experience of decision makers, who evaluate multiple options in order to choose between them (e.g., marketing managers; Shen et al. 2012). In particular, the *joint-evaluation* design used in the expert survey is consistent with how fundraisers evaluate the advantages and disadvantages of the fundraising strategies under consideration. By comparing multiple options, the managers can prioritize among strategies or select different strategies for different target groups.

Experts saw five appeals (see Table 1) and compared four pairs of appeals. Specifically, participants contrasted standard matching with the control, standard matching with a giving-credit framing with standard matching alone, threshold matching with standard matching, and threshold matching with a giving-credit framing with threshold matching alone.

For each pair, participants first evaluated the likelihood that participation rates (i.e., the number of people responding to the appeal) would be higher in one appeal versus the other. Then they evaluated how likely it was that the average donation amount (i.e., among those who responded to the appeal with a nonzero donation) would be higher in one appeal versus the other. For each of these questions, responses were captured using a five-point Likert scale (1 = definitely yes; 2 = probably yes; 3 = cannot predict; 4 = probably no; and 5 = definitely no) comparing the efficacy of one appeal with the other in the pair. Last, a few follow-up questions were asked, including about their work experience, and an instructional manipulation check (Oppenheimer et al. 2009) was performed that participants needed to answer correctly to finish the survey.

3.2. Results

The raw distribution of experts' responses is shown in Figure 1. Our analysis was intended to determine whether experts were significantly more likely to evaluate one of the solicitation variations as performing better than the other. Accordingly, we recoded responses "definitely" or "probably" favoring one appeal on the Likert scale as +2 and +1, respectively; neutral responses of "cannot predict" as 0; and responses "probably" or "definitely" favoring the other

Figure 1. Distribution of Expert Responses in Study 1

Notes. The bars indicate the percentages of experts agreeing to a particular rating point. The charts on the left indicate participation, whereas those on the right indicate average contributions. GC, giving credit.

appeal as -1 and -2 , respectively. We then compared the mean response of the recoded values to zero.

3.2.1. Predicted Participation. Comparing standard matching with the control, 90% of the experts thought participation would be higher (mean (M) = 1.22 versus 0, $t(104) = 15.45$, $p < 0.001$) with standard matching, and only 6% thought participation would be lower.¹ This overwhelmingly positive expectation is at odds with the mixed empirical evidence in prior research. The consensus that matching improves participation could reflect experts' beliefs that matching would not crowd out donations in the altruism model or could mean that experts relied on perceptions that their own experience with matching had been positive and did not consider the potential crowding-out effect of matching.

More relevant to our research question, the experts further expected that the giving-credit framing would boost participation for matching solicitations. Comparing standard matching with the giving-credit framing, 70% of the experts thought the giving-credit framing would yield higher participation, consistent with the model predictions we derived earlier, whereas only 14% thought participation would be lower ($M = 0.74$ versus 0, $t(104) = 7.92$, $p < 0.001$).

However, the experts were more pessimistic about threshold matching. Experts were split on whether such conditional matching would yield higher participation than standard matching, with 35% saying that it would be higher and 40% saying that it would be lower ($M = 0.02$ versus 0, $t(104) = 0.17$, $p = 0.867$). This is consistent with the lack of clear predictions for threshold matching we identified in the impure altruism model and the need for additional strong assumptions, which could vary across experts, in order to make a determination. Nevertheless, even in threshold matching offers, the experts were more likely to believe that the giving-credit framing would increase participation relative to the standard framing (53%) rather than lower participation (19%; $M = 0.39$ versus 0, $t(104) = 4.23$, $p < 0.001$), consistent with the model predictions.

3.2.2. Predicted Contribution Size. Overall, expert predictions regarding contribution size revealed patterns similar to those of their predictions for participation. Comparing standard matching with the control, 63% of the experts thought that contributions would be larger with standard matching, whereas only 17% thought that contributions would be smaller ($M = 0.67$ versus 0, $t(104) = 6.53$, $p < 0.001$). A majority of experts also thought that the giving-credit framing would increase contribution amounts, with 65% responding that matching with the giving-credit framing would yield larger contributions than regular fram-

ing under standard matching, whereas only 15% thought that they would be smaller ($M = 0.64$ versus 0, $t(104) = 6.94$, $p < 0.001$).

Experts were split but overall less optimistic about the effects of threshold matching on the amounts donors gave. Only 49% of the experts thought that threshold matching would yield larger contributions than standard matching, and 32% thought that it would result in smaller contributions ($M = 0.21$ versus 0, $t(104) = 1.97$, $p = 0.051$). Nevertheless, a majority of experts predicted that the giving-credit framing would yield larger donations than regular framing even in a threshold matching appeal (57%), and few believed the reverse (17%; $M = 0.45$ versus 0, $t(104) = 4.93$, $p \leq 0.001$).

Overall, the results suggest that experts strongly agreed with the predictions of the impure altruism model: the giving-credit framing would outperform regular framing whether using a standard or threshold match, but their predictions did not reflect consensus about the success of threshold matching versus standard matching.

As discussed earlier, a simultaneous joint evaluation of potential fundraising strategies closely represents how such ideas are evaluated in practice. However, this raises the possibility that experts could mispredict the effects of interventions on donor behavior specifically because the donors, unlike the experts, evaluate only a single appeal in isolation (Shen et al. 2012). If this is the case, expert fundraisers could make better decisions by splitting the task and examining one idea per person (or per group) before coming together to decide on the best approach.

To test for the robustness of our findings to experts' evaluation mode, we conducted another study in which we randomly assigned experts to first evaluate one of two fundraising appeals (separate evaluation) and only then introduced and had them evaluate the other appeal (allowing joint-evaluation tests). We also asked follow-up questions about the perceived effects of the appeals on potential donors' attitudes and behavior. This was done to more directly examine the extent to whether experts' reasoning corresponded with the components of the impure altruism model.

4. Study 2: Robustness of Expert Opinion to Evaluation Mode

4.1. Design

We used the services of the same professional online panel company to recruit $n = 200$ managers responsible for designing or implementing fundraising strategies. Participants had an average of 10 years of experience in fundraising-related work, and 69% reported having worked in fundraising campaigns that specifically used matching contributions.

In this survey, we compared only standard one-to-one matching with and without the giving-credit framing (i.e., we did not test threshold matching) to increase statistical power for examining differences in separate evaluations. Specifically, experts were first introduced to one of these two appeals (counter-balanced) and asked to rate the appeal on several measures. Then, on a separate page without prior notice, experts were asked to examine the other appeal, which they also rated on the same measures. The details shared with the participants about the appeals, the nature of the charitable organization soliciting funds, the type of donors being solicited (i.e., prior donors only), and the mode of solicitation (i.e., postal mail) were similar to those in the previous survey (see the online appendix).

Using balanced six-point scales, participants first rated each appeal on its likelihood to increase participation and donation size. Next, on a separate page, experts reported their beliefs about the three main constructs of the impure altruism model in lay terminology: potential donors' feelings about the overall donation opportunity, warm-glow feelings, and perceived altruistic benefit (the opportunity to help others). Finally, on the same page, we elicited experts' beliefs about a few potential psychological determinants of donors' behavior: donors' feeling of personal responsibility for the funds going to the charity, as well as the appeals' ease of understanding and perceived coerciveness (see the online appendix for full details of the stimuli used).

The appeal being evaluated was always displayed on the page where the participants answered these questions so that the experts did not have to rely on their memory. When experts evaluated the second appeal, both appeals were presented on the page to facilitate comparison. Furthermore, experts were reminded about how they had rated the same question for the first appeal when they evaluated the second. This was again done to make sure that experts did not have to rely on their potentially incorrect memory when comparing the two fundraising appeals.

A few follow-up questions were also asked, including questions on work experience and an instructional manipulation check (Oppenheimer et al. 2009) that participants needed to complete correctly to finish the survey.

4.2. Results

4.2.1. Between-Subject Evaluations. Experts rated the giving-credit framing as similar to standard matching in the likelihood of increasing participation ($M_{\text{giving credit}} = 4.17$ versus $M_{\text{standard matching}} = 3.99$, $t(198) = 1.11$, $p = 0.268$). Likewise, experts did not predict a higher average contribution among participating donors for the giving-credit framing ($M_{\text{giving credit}} = 3.97$

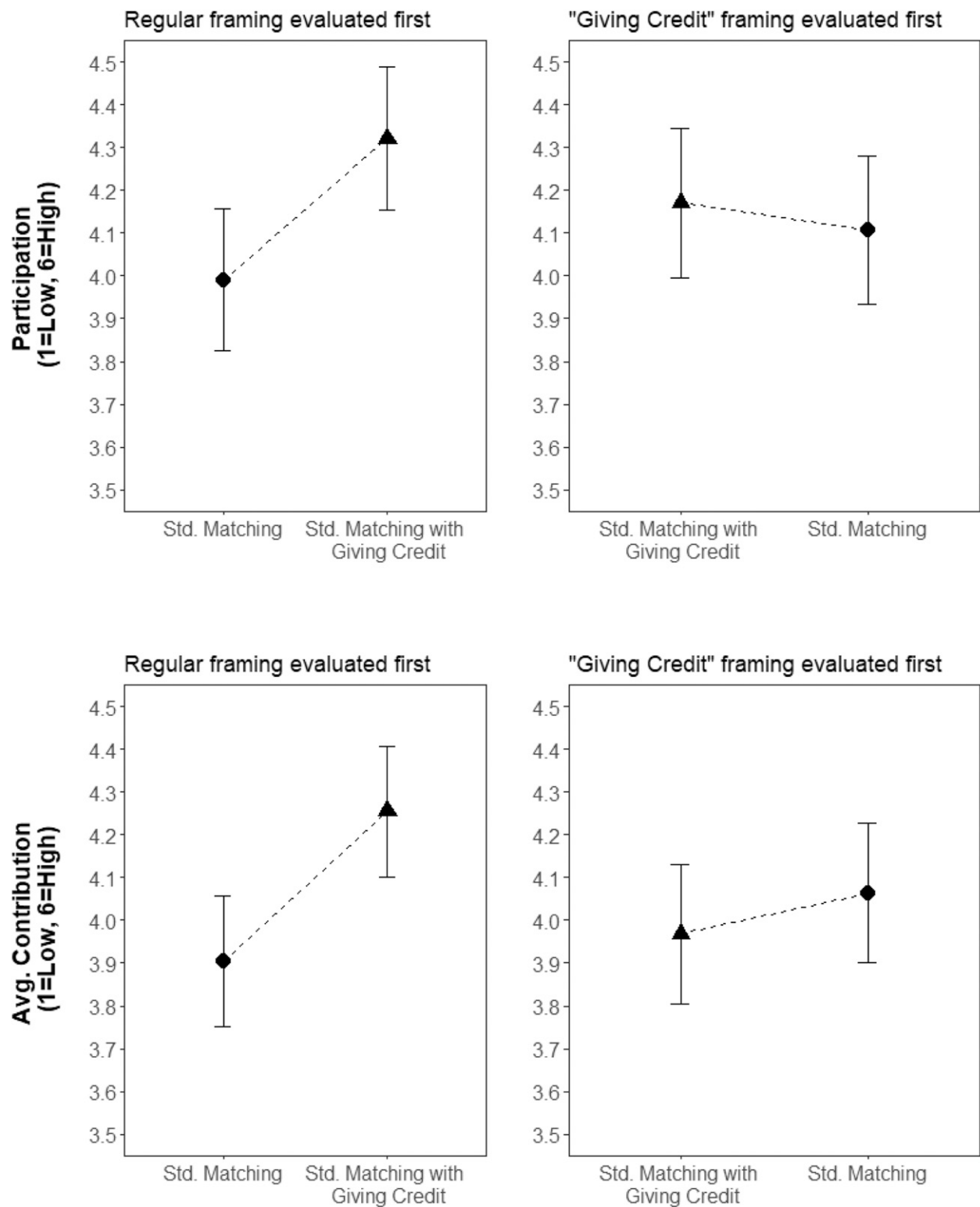
versus $M_{\text{standard matching}} = 3.90$, $t(198) = 0.40$, $p = 0.688$). Therefore, experts did not perceive the overall efficacy of the two match framings differently when evaluating them separately. This is in contrast to the overwhelmingly more positive evaluation experts had for the giving-credit framing in Study 1, where they directly compared the two appeals.

Experts did perceive the giving-credit framing as impacting the constructs in the impure altruism model, causing significantly higher feelings of warm glow among donors ($M_{\text{giving credit}} = 4.57$ versus $M_{\text{standard matching}} = 4.18$, $t(198) = 2.16$, $p = 0.032$) and significantly higher perceptions of a good opportunity to help others ($M_{\text{giving credit}} = 4.70$ versus $M_{\text{standard matching}} = 4.35$, $t(198) = 2.09$, $p = 0.038$). Both of these effects were completely mediated by a marginally significant indirect effect via higher expected positive feelings of personal responsibility for the funds raised by the charity (95% bootstrapped confidence intervals (CIs) of the indirect effect = $[-0.03, 0.46]$ and $[-0.03, 0.46]$, respectively; see the online appendix for additional details). Experts did not judge the two appeals differently in terms of ease of understanding ($p = 0.731$) or the extent to which donors might perceive the appeals as pushy or manipulative ($p = 0.603$).

Overall, in separate evaluations, the standard match and the standard match with the giving-credit framing appeal were seen as similarly effective by the experts. However, even the separate-evaluation expert judgments validated the model assumptions that potential donors were more likely to feel good about giving when a standard match was framed to give them credit for the funds received. The evidence suggested that this was on account of experts' beliefs that donors might feel more responsible for the funds raised by the charity when the appeal was framed to give donors the entire credit for the donations (i.e., including the matched amounts).

4.2.2. Within-Subject Evaluations. Comparing the experts' ratings across the two potential appeals each expert rated (order counterbalanced) reveals a significantly more favorable view of the giving-credit framing compared with the standard framing. This was particularly the case when the framing was evaluated *after* first rating the standard matching appeal (see Figure 2). Experts predicted that the giving-credit framing would significantly increase participation compared with standard matching alone ($M_{\text{giving credit}} = 4.25$ versus $M_{\text{standard matching}} = 4.04$, $t(199) = 2.38$, $p = 0.018$) and would marginally increase the average dollar contribution from participating donors ($M_{\text{giving credit}} = 4.12$ versus $M_{\text{standard matching}} = 3.98$, $t(199) = 1.73$, $p = 0.085$). In particular, the predictions favored the giving-credit framing when this novel appeal was evaluated after the standard

Figure 2. Expert Ratings of the Ability to Increase Participation (Upper Panels) and Nonzero Contributions (Lower Panels) in Study 2, Measured on a One-to-Six Scale



Notes. The charts are broken by the order in which the appeals were evaluated. The vertical lines are 95% confidence intervals (corrected for repeated measures). Std., standard.

matching appeal (participation: $M_{\text{giving credit}} = 4.32$ versus $M_{\text{standard matching}} = 3.99$, $t(105) = 2.77$, $p = 0.006$; interaction with order of evaluation: $\beta = 0.26$, $t = 1.55$, $p = 0.122$; average contribution: $M_{\text{giving credit}} = 4.25$ versus $M_{\text{standard matching}} = 3.90$, $t(105) = 3.20$, $p = 0.002$; interaction with order of evaluation: $\beta = 0.44$, $t = 2.79$, $p = 0.005$). The experts' joint-evaluation predictions replicate the results of the prior survey (Study 1) using a different research design and elicitation approach.

Experts in joint evaluation also rated the giving-credit framing as significantly better at increasing donors' feelings of warm glow ($M_{\text{giving credit}} = 4.58$ versus $M_{\text{standard matching}} = 4.30$, $t(199) = 3.63$, $p < 0.001$). Consistent with the assumptions we made in the model-based predictions derived earlier, experts' higher warm-glow predictions mediated their more positive evaluation of the giving-credit framing appeal in terms of both participation (indirect effect 95% CI = [0.08, 0.28]) and donation size (indirect effect 95% CI = [0.08, 0.26]).

Furthermore, compared with the standard matching appeal, experts believed that potential donors' would perceive the giving-credit framing as providing a better opportunity to help others (i.e., a measure of pure altruism; $M_{\text{giving credit}} = 4.67$ versus $M_{\text{standard matching}} = 4.44$, $t(199) = 3.38$, $p < 0.001$). Experts' higher judgment of the pure altruism benefits also mediated their more positive evaluation of the giving-credit framing appeal in terms of both participation (indirect effect 95% CI = [0.06, 0.24]) and donation size (indirect effect 95% CI = [0.06, 0.22]).

Finally, experts' overall joint-evaluation judgments indicated that the giving-credit framing would be perceived by donors as a more beneficial opportunity to donate (i.e., a measure of perceived total utility from giving; $M_{\text{giving credit}} = 4.69$ versus $M_{\text{standard matching}} = 4.52$, $t(199) = 2.33$, $p = 0.020$). Furthermore, the higher perceived overall benefits mediated experts' more positive evaluation of the giving-credit framing appeal in terms of both expected participation (indirect effect 95% CI = [0.02, 0.21]) and expected donation size (indirect effect 95% CI = [0.02, 0.18]).

Experts also predicted that donors would feel more responsible for the money raised through the giving-credit appeal versus the standard matching appeal ($M_{\text{giving credit}} = 4.35$ versus $M_{\text{standard matching}} = 4.00$, $t(199) = 4.42$, $p < 0.001$). Expectations about donors' greater feeling of personal responsibility mediated experts' more favorable judgment of the giving-credit framing over a standard matching appeal in terms of both participation (indirect effect 95% CI = [0.12, 0.30]) and donation size (indirect effect 95% CI = [0.11, 0.29]). Furthermore, as was observed for separate evaluation, the positive feeling of personal responsibility for the funds raised in the giving-credit framing partially

mediated the effect of the giving-credit framing on generating higher feelings of warm glow among donors (95% bootstrapped CI of the indirect effect = [0.09, 0.30]) and completely mediated the effect of the framing on higher perceptions as a good opportunity to help others (95% bootstrapped CI of the indirect effect = [0.13, 0.33]).

Even when an appeal is seen as potentially more effective in theory, fundraisers may be reluctant to use it based on concerns about unintended negative consequences in practice. However, experts did not perceive the giving-credit framing as more pushy or manipulative and in fact actually found it less so than the standard matching appeal ($M_{\text{giving credit}} = 2.70$ versus $M_{\text{standard matching}} = 2.86$, $t(199) = 2.10$, $p = 0.037$). Experts also did not anticipate that donors would find the giving-credit appeal more difficult to understand than the standard matching appeal ($M_{\text{giving credit}} = 4.74$ versus $M_{\text{standard matching}} = 4.75$, $t(199) < 1$).

4.3. Discussion

Overall, these results suggest that experts' predictions (particularly in joint evaluation) parallel our assumptions in extending the theoretical model of impure altruism. Experts expected donors viewing a giving-credit appeal to feel more responsible for the funds raised. As developed in our extended impure altruism model, experts then predicted that the giving-credit framing would yield significantly more warm-glow feelings among potential donors and that they would perceive it as a better opportunity to contribute. These beliefs, in turn, mediated their overall impressions—consistent with the model—that the giving-credit framing would result in better fundraising outcomes than the standard matching appeal.

However, the experts' reasoning potentially violated the all-else-equal assumption, in that they also perceived the giving-credit framing as increasing purely altruistic benefit to the donor. Furthermore, experts' predictions were affected by both the evaluation mode and order. Their efficacy predictions were less conclusive in separate evaluation and in joint evaluation when rating the giving-credit framing appeal first. Further analyses (see the online appendix) suggest that order effects in how experts perceived donor utility from the two appeals may account for the order effects in experts' efficacy predictions.

Last, we briefly mention one other potential source of guidance other than academic models and expert opinion. Survey research (i.e., of potential donors) is rarely used in fundraising, and when it is used, it tends to focus on donor attitudes toward nonprofits rather than testing fundraising tactics.² However, surveys are a widely used approach to predicting outcomes in other domains (e.g., voting, consumer decisions). For completeness, we also tested the fundraising

tactics using a survey of laypeople (e.g., potential retail donors), with each participant responding to one appeal (i.e., separate evaluation). The results are inconclusive, similar to the separate-evaluation expert data, with no significant evidence that the giving-credit-framed appeal would perform better but also no indication that using the framing would negatively impact fundraising and no significant differences from the threshold match (see the online appendix for the details of this additional study).

Overall, the sources of information discussed up to this point would suggest to the fundraiser that using the giving-credit framing has little risk and is likely to yield better results than the standard match, particularly if jointly evaluating more than one intervention when making the decision. These sources of information would suggest more uncertainty about the effect of the threshold match, failing to yield a single consensus prediction. How sufficient and accurate would these sources of guidance be for a diligent fundraiser trying to employ best practices to decide on the final campaign design? To assess this question, we conducted a fundraising field experiment in collaboration with a community-based nonprofit in Chicago.

5. Study 3: Field Experiment to Evaluate Guidance

5.1. Design

The nonprofit conducting the fundraising campaign was a small but well-established arts organization with fewer than 15 employees and a long history in the community. It promotes young artists by organizing exhibitions and workshops and also offers various art classes, including summer art classes for children. The nonprofit was planning to conduct its annual fundraising campaign and, as part of celebrating its 75th anniversary, had secured a leadership gift from one of its patrons for that year's campaign. The organization leveraged the leadership grant to run a matching campaign during this fundraising drive, with five different randomized mail-based solicitations. This experiment, unlike the survey experiment, represents a natural field experiment (Harrison and List 2004) in that decision makers did not know that the fundraising appeals were part of an experiment.

The organization sent out mailers with a letter signed by the deputy director, a pledge card, and a prepaid self-addressed envelope. Mailers were sent to the organization's list of 3,588 potential retail donors. The experimentally manipulated matching offers, however, were sent only to the people who had donated previously, which constituted 1,480 mailings. Targeting prior donors is common in fundraising field studies that use mailers (List and Lucking-Reiley 2002; Karlan and

List 2007, 2012; Huck and Rasul 2011; Karlan et al. 2011; Goswami and Urminsky 2016a) because of substantially higher participation rates, potentially due in part to a greater willingness to open and read the solicitation. The targeted prior donors were primarily small-amount contributors (median last contribution of \$45) who had previously bought a membership, enrolled in a class, attended an event, or contributed in some other way to the organization.

The matching offer for each experimental condition, when applicable, was presented both in the body of the letter and in a summary of the matching offer printed on the back flap of the self-addressed envelope (see the online appendix for full details of the stimuli used). The prior donors were each sent one of the five experimental mailers, using a 2 (matching mechanism: one-to-one standard versus threshold) \times 2 (framing: regular versus giving credit) $+ 1$ (no-match control) between-subjects randomized design. Mailers were sent out in the first week of September 2014. Contributions were recorded until February 2015 (i.e., for about five months), by which time contributions to the campaign had largely ended (only two contributions were received in February).

5.2. Results

We analyzed three outcomes: participation, average contribution among donors (i.e., conditional on sending in a donation), and net money raised, as in the experts' surveys.

5.2.1. Actual Participation. Averaging across experimental conditions, the overall contribution rate was 5.6%.³ The participation level in the standard one-to-one matching condition (8.1%) was directionally higher than in the control condition (5.1%), but the difference was not significant ($\chi^2(1) = 2.25, p = 0.133$). This result is consistent with the mixed results in the prior literature on the effects of using matching solicitations with prior donors but inconsistent with the near-consensus opinion among experts in Study 1, 90% of whom thought that matching would be effective in increasing participation.

More important, we find no evidence that the giving-credit framing improved donation rates, and the observed difference instead points in the opposite direction. Contrary to both the predictions of the extended impure altruism model and the expert practitioners, when the solicitation was framed to give donors credit for the matching funds, the participation rate was *lower* than with the regular framing, significantly under the standard match (4.1% versus 8.1%; $\chi^2(1) = 4.06, p = 0.043$) and directionally under the threshold match (3.9% versus 6.6%; $\chi^2(1) = 2.08, p = 0.148$).

Overall, collapsing across standard and threshold matches, the giving-credit framing significantly reduced participation (4.0% versus 7.4%; $\chi^2(1) = 6.21$, $p = 0.013$). The detrimental effect of the giving-credit framing did not vary depending on the matching mechanism used (standard versus threshold: $\beta = 0.149$, $z = 0.281$, $p = 0.778$).

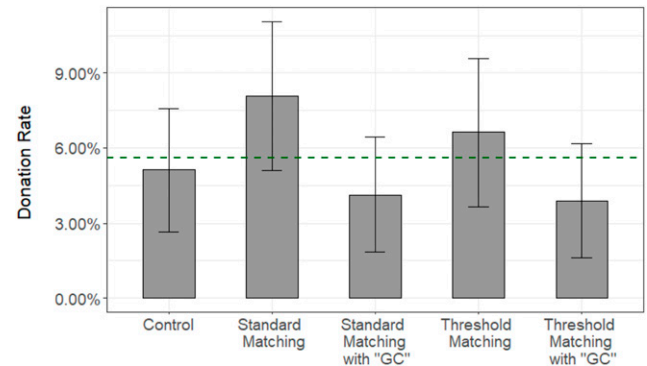
The effects of threshold (versus standard) matching did not diverge from the model and expert predictions but were inconsistent with recent academic endorsements of threshold-based matching (Sanders et al. 2013). We found no detectable improvements in participation using a threshold match versus a full match, either under the regular framing (6.6% versus 8.1%; $\chi^2(1) = 0.45$, $p = 0.499$) or under the giving-credit framing (3.9% versus 4.1%; $\chi^2(1) = 0.02$, $p = 0.878$). Consequently, across the various framing manipulations, there was no overall difference in participation for the different types of matching mechanisms (5.2% versus 6.2%; $\chi^2(1) = 0.52$, $p = 0.471$), consistent with the equivocal implications of the model (absent additional assumptions) and with the mixed predictions of the experts.

Overall, although both sources of guidance may have correctly captured the various levers affecting prosocial motivation, they failed to predict the extent and direction of how warm-glow motivation might be affected by solicitation framing, at least in this case. Consequently, the effect of the giving-credit framing on participation was favorably estimated by both sources, whereas empirical evidence found a significantly detrimental effect of this framing on participation, particularly when compared with standard matching. However, expert and model predictions resonated with the empirical results that found an ambiguous effect of the threshold matching mechanism over the standard matching mechanism. Finally, whereas experts were overwhelmingly positive about the effect of a standard matching appeal over a no-match appeal, the empirical results were, at best, directional, consistent with the mixed implications from the model.

5.2.2. Actual Contribution Size. Next, using log transformations to account for the skew in the donation amounts, we calculated the average donation size per condition (see Figure 4).⁴

Those prior donors who gave in this campaign contributed directionally less in the standard one-to-one match than in the control condition ($t(40) = 1.34$, $p = 0.185$; see Figure 3). Again this highlights the double-edged nature of standard matching, which is consistent with some interpretations of the theoretical model and illustrates the necessity of making additional

Figure 3. Participation in Various Experimental Conditions in Study 3



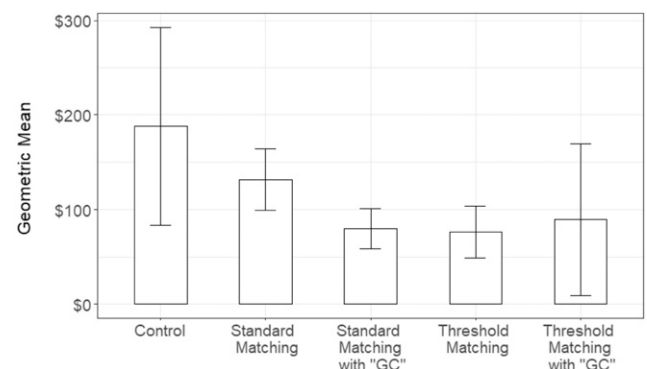
Notes. The overall participation rate of 5.6% is indicated by the dotted horizontal line. The vertical lines are 95% confidence intervals. GC, giving credit.

assumptions to use the impure altruism model for policy predictions.

More important, we find no evidence that the giving-credit framing increased contribution amounts. In fact, the giving-credit framing significantly reduced contributions relative to regular framing for the standard one-to-one match offer ($t(36) = 2.43$, $p = 0.020$) and resulted in similar donations under threshold matching ($t(27) < 1$). The effect of framing (regular versus giving credit) did not significantly differ by match type (standard one-to-one versus threshold matching: $\beta = 0.650$, $t = 1.57$, $p = 0.121$). Therefore, both model-based and joint-evaluation expert guidance incorrectly predicted the positive effects of the giving-credit match, and even experts in the separate-evaluation mode failed to correctly anticipate the negative effects.

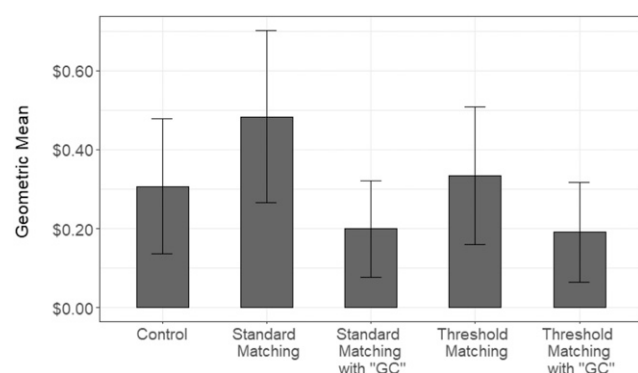
The goal of the threshold match was specifically to reduce the crowding-out effects of a standard match and thereby yield larger donations among those who give. The threshold matching intervention failed in

Figure 4. Average Contribution Among Donors in Study 3



Notes. The vertical lines are 95% confidence intervals. GC, Giving credit.

Figure 5. Net Donations per Mailing in Study 3



Notes. The vertical lines are 95% confidence intervals. GC, giving credit.

this regard, yielding significantly smaller contributions than the standard match ($t(42) = 2.59, p = 0.013$) under the regular framing and no difference under the giving-credit framing ($t(21) < 1$). Collapsing across different types of framing, the threshold match yielded directionally lower contributions than a standard matching solicitation ($t(65) = 1.63, p = 0.107$) among those choosing to donate. This directional but inconclusive result is somewhat consistent with the lack of clear prediction from the model and with the inconsistent predictions of the experts.

Overall, the results for donation size largely paralleled the results for participation. The empirical results of the giving-credit framing systematically diverged from the predictions, whereas the lack of difference between the standard match and the threshold match aligned with the lack of clear model implications and with the lack of expert consensus. Finally, the empirical results of the standard matching appeal over an appeal without any matching on donation amounts were more consistent with the ambiguous implications of the model and diverged from the positive opinions of the experts (63% reported an improvement in donation outcomes).

5.2.3. Actual Net Money Raised. To assess the net effects, incorporating both the number of donors and how much each donor gave, we compared the log-transformed average money raised per mailing (including zero contributions) across conditions. On account of directionally higher participation, the standard one-to-one match raised directionally more money than the control no-match condition ($t(633) = 1.26, p = 0.208$).

However, we find no evidence that more money is raised when using the giving-credit framing than regular framing. In fact, the giving-credit framing raised significantly *less* money per person under standard matching ($t(610) = 2.28, p = 0.022$) and

directionally less under threshold matching ($t(553) = 1.32, p = 0.188$). The effect of framing (regular versus giving credit) did not significantly differ by match type (standard one-to-one versus threshold matching: $\beta = 0.100, t = 0.78, p = 0.431$). Overall, collapsing across standard and threshold matching, the giving-credit framing significantly reduced the money raised per mailing compared with the regular framing ($t(1165) = 2.63, p = 0.008$; see Figure 5). To the degree that net money raised or the number of donors participating is the primary metric of concern to a fundraiser, these results would be disappointing for the fundraiser, particularly given the expectations set by the prior guidance.

Combining the effect on participation and the average contribution, there was no significant effect of threshold (versus standard) matching on net money raised under either the regular framing ($t(592) = 1.05, p = 0.295$) or the giving-credit framing ($t(571) < 1$). Combining across the framing manipulations, there was no overall difference in net effect for the different types of matching mechanisms ($t(1165) < 1$). This suggests that the theorized potential of a threshold matching mechanism to arrest the crowding-out effects of a standard matching mechanism (Sanders et al. 2013) did not occur in practice. However, at the same time, the lack of a difference is not inconsistent with the equivocal implications of the model and with the disagreement among the experts.

5.3. Discussion

The results of the field study are at odds with some of the expectations a fundraiser would have from both academic and industry expert sources of guidance. Both the impure altruism model and the majority of expert practitioners predicted better fundraising performance when using the giving-credit framing. However, the giving-credit framing not only failed to improve fundraising but instead consistently reduced the outcomes. In addition, although some academics have advocated for threshold matching, the impure altruism model yields inconclusive predictions (without strong additional assumptions), and experts were split on whether it would be more effective.

It is noteworthy how few experts predicted the outcome of this study. Only 12% in Study 1 and 18% in Study 2 predicted that the giving-credit framing would do worse than regular framing, both using standard matching, and only 9% in Study 1 predicted that threshold matching would not make a difference compared with standard matching, both using regular framing. In fact, none of the 105 experts in Study 1 correctly predicted the effects of both the giving-credit framing and the threshold match.

The results of the field experiment cannot be explained as a failed manipulation (i.e., donors not being affected by the framing). There is a consistent decrease with the giving-credit framing compared with the regular framing used in standard matches across the three key metrics: participation, average contribution, and net money raised. Furthermore, threshold matching with the giving-credit framing was directionally worse than threshold matching alone for both participation and net money raised. Such consistency in results would be unlikely if potential donors did not notice the experimental manipulation.

Nevertheless, the results are not extremely strong, and the multiple tests involved in our analyses increase the risk of false-positive results. With the benefit of hindsight, the five-condition study was underpowered to detect the size of effects when comparing across conditions that were observed. Therefore, before we can definitively conclude that this field study established a ground truth contrary to the recommendations from theoretical models and from experts, it would be useful to conduct a higher-powered confirmatory replication test with preregistered comparisons.

To do so, we again partnered with the same organization, in the spring of 2018, to conduct a new fundraising experiment, three and a half years after the first study. Given that the number of prior donors constrains the sample size, the most feasible approach to increase statistical power was to reduce the number of conditions. Thus, in this new study, we focused the research design on one key comparison: a giving-credit framing versus regular framing for a standard one-to-one match.

6. Study 4: Field Replication as a Test of Generalizability Across Campaigns

6.1. Design

The nonprofit organization generally runs its annual campaign in the fall of every year, but it agreed to run an additional campaign, in which we implemented our two-condition experiment, in the spring. The study and the analysis plan were preregistered (viewable at <http://aspredicted.org/blind.php?x=na3yk9>). Several aspects of the design were similar to the previous field experiment. First, participants were not aware that they were taking part in an experiment. Second, mailers were sent out with a letter signed by the deputy director, a pledge card, and a prepaid self-addressed envelope. The experimental intervention was implemented both in the letter and on the self-addressed envelope.

However, there were a few differences from the previous field experiment. Mailers were sent to 3,646 people, of whom 3,036 were prior donors, and the remaining 610 were nondonors. Prior donors who had responded to the fall 2017 campaign and had

donated more than \$250 were excluded from the campaign. Likewise, donors for whom the organization had other upcoming fundraising plans, typically individually customized, were excluded prior to randomization. The median last contribution from prior donors in the new study was \$190, which was higher than in the previous study.

Nondonors were people who had never donated to the organization but had attended a free event hosted by the organization in the past five years. The sample was randomly divided into two experimental cells: one group received a standard one-to-one matching solicitation, and the other received a one-to-one matching solicitation with giving-credit framing (see the online appendix for details). Note that whereas the threshold matching scheme precluded the use of nondonors in Study 3, the use of only standard matching in this experiment (with either the regular framing or the giving-credit framing) allows us to potentially include prospective donors.

All letters were sent out in the second week of May 2018. We collected responses until the end of the second week of August 2018—a period of over three months from the start of the campaign.

6.2. Results

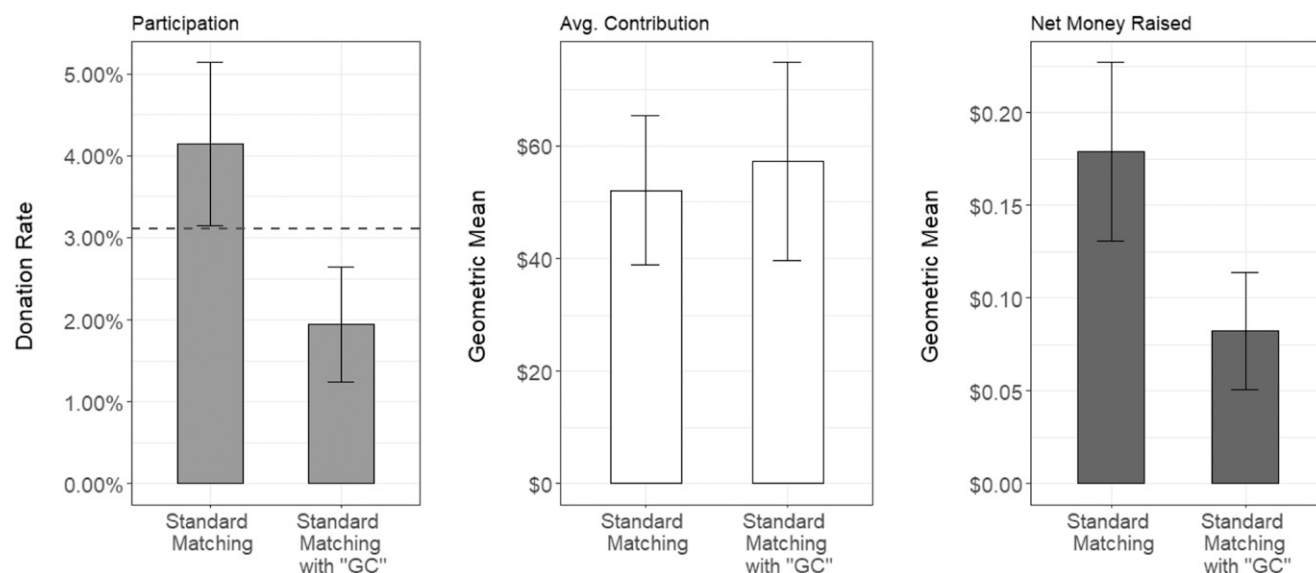
There was only one donation (of \$50) received from a nondonor (who had responded to the standard one-to-one matching solicitation). Therefore, the following analysis considers data only from the prior donors, although the conclusions do not differ if the nondonors are included.

Averaging across experimental conditions, the overall contribution rate was 3.1%.⁵ However, replicating the prior field experiment, the participation level in the standard one-to-one matching condition (4.1%; see Figure 6, first panel) was significantly higher than in the giving-credit framing condition (1.9%; $\chi^2(1) = 12.34, p < 0.001$).

As in the original field experiment, we used log-transformed donation amounts to account for skew in the data. As shown in the middle panel of Figure 6, the average donation size among those who gave did not significantly differ between the two conditions ($t(91) < 1$). In the previous field study, we had found that the giving-credit framing yielded significantly smaller contributions among participating donors when the standard one-to-one match was used. Although we do not replicate this finding, we again find no evidence that the giving-credit framing *increases* the size of donations, as predicted by the theoretical model and by expert fundraisers.

Overall, the standard one-to-one match raised significantly more money per mailing than the giving-credit framing ($t(3,034) = 3.33, p < 0.001$), driven by the difference in donation rates. Therefore, in this

Figure 6. Participation, Average Contribution Among Donors, and Net Donations per Mailing in Study 4



Notes. The overall participation rate of 3.1% is indicated with a dotted horizontal line in the left-hand chart. The vertical lines are 95% confidence intervals. Avg., average; GC, giving credit.

preregistered and higher-powered field replication, we confirm that the giving-credit framing was significantly less effective than the regular framing of the matching solicitation. These results are robust to using non-parametric statistical tests, excluding outliers, or controlling for covariates.⁶

6.3. Discussion

In the replication study, a larger sample size and preregistration were employed to conduct confirmatory tests of the giving-credit findings in Study 3. The lower performance of the giving-credit framing was replicated in the same fundraising context, driven by the replicated difference in donation rates.

7. General Discussion

In this paper, we explore the adequacy of the sources of guidance available to a professional fundraiser choosing whether to implement a novel form of matching solicitation involving either alternative framing or a customized minimum threshold for matching. We compare the guidance the fundraiser could receive from two sources: (1) an extended version of the theoretical impure altruism model of donation behavior and (2) the opinions of expert fundraisers. In general, these sources of guidance predicted that the giving-credit manipulation would either have no effect (if people failed to incorporate the match into their warm-glow utility) or a positive effect and did not make a clear prediction for threshold matching.

Given the common reluctance of fundraisers to conduct field experiments, our hypothetical fundraiser might well decide to implement the giving-credit framing

(with or without a threshold match) based on one or more of these sources of guidance. The results of two expert surveys, including one where the evaluation modality strongly mimicked how donors might evaluate an appeal, did not highlight any potential downsides of the new framing appeal. The results of our field experiment suggest that this decision, reasonably based on the likely sources of information available to the fundraiser, might be a serious mistake in the current fundraising context. The results of both the original field experiment and a follow-up preregistered replication field experiment consistently found strong evidence that the giving-credit framing would result in significantly worse outcomes for the fundraising organization.

It is important to note that the ground truth identified in our field experiments might be context specific. Our studies were conducted in one particular fundraising setting, with one donor pool, and we cannot know the degree to which these findings generalize across fundraising campaigns. In fact, research has repeatedly found that the effects of behavioral interventions vary by field context (Alcott 2015). For example, although Landry et al. (2006) found a positive effect of solicitor–solicitee interaction in a door-to-door campaign, a subsequent study (List and Price 2009) that matched solicitors and solicitees in terms of race and gender in a door-to-door fundraising drive failed to improve the performance of the campaign. Our findings highlight the complexity of the psychology of the warm-glow motivation and, consequently, the limitations of model-based predictions and expert forecasts for novel situations. Thus, our findings demonstrate the importance of not

only field experiments in general but also sufficiently powered *in-context* field experiments.

7.1. Implications for Research on Fundraising

Why did the seemingly promising giving-credit framing of fundraising appeals fail in the field? It appears that potential donors may have failed to incorporate the intended benefits into their warm-glow utility. Warm-glow preferences were first proposed to explain empirical anomalies that were difficult to reconcile with a theory based on pure altruism (Andreoni 1988, 1990). Notwithstanding the improvements this new theory afforded to our understanding of altruistic behavior, our results suggest that more model development and empirical research on which factors facilitate or inhibit warm-glow preferences are needed before we can reliably predict the effects of novel fundraising interventions on donation behavior.

One possibility is that the giving-credit framing was seen as mixing egoistic benefits (i.e., benefits to one's self) with altruistic benefits (i.e., benefits to others), which some research has suggested reduces the performance of appeals, compared with only one or the other benefit (Feiler et al. 2012, Dubé et al. 2017). The researchers have suggested that this happens because messages employing mixed motives are construed as an overt attempt to persuade, which might then be seen as a coercive nudge (e.g., Fitzsimons and Lehmann 2004), inhibiting warm glow. Although we do not know if that is the case in our field studies, it is notable that experts' opinions elicited in Study 2 on specifically this possibility did not reveal a concern that the giving-credit manipulation would be seen as manipulative.

Another possibility is a generalization of intuitions that underlie social loafing (Latané et al. 1979). By telling potential donors that the leader would add to specifically their own donations, the potential donors might have felt less accountable for their own donations and a low sense of ownership of the resulting "joint" donations, undercutting their motivation to donate. The experts failed to anticipate this possibility and, in Study 2, made the exact opposite prediction, that the giving-credit framing would increase feelings of responsibility for the funds raised.

It is also possible that the giving-credit framing was more cognitively demanding to process, and potential donors therefore deferred their decisions, resulting in a lower donation rate. However, to the degree that the response of the potential donors to the giving-credit framing reflects any of these impressions of the giving-credit appeals, neither our extension of the impure altruism model nor the fundraising experts anticipated them. In fact, the experts in Study 2 rated both appeals as equally easy or difficult to understand.

Of course, our findings in no way invalidate the impure altruism model either but instead illustrate its

limitations when used to make decisions. The impure altruism modeling framework is flexible enough to accommodate all of these possibilities given additional assumptions or parameters. However, although such a modeling approach could describe the results post hoc, it would not provide the kind of ex ante model predictions that are needed as a source of guidance to evaluate the viability of novel fundraising solicitations. It might be reasonable to expect experts, using reasoning based on the model and filling in the missing assumptions based on their experience, to provide better guidance. However, although expert reasoning did closely parallel the constructs in the model, experts were not able to consistently predict the outcomes tested.

By highlighting the incompleteness of the theory of warm glow, our findings raise important questions that would need to be addressed in future research to develop analytical models that can make more precise predictions in the domain of fundraising. Perhaps these findings may motivate research that moves from the field back to the lab such that nonpredicted findings in the field motivate more basic research in laboratory settings (i.e., as suggested by Bartels et al. 2018). However, this would require capturing the key psychological factors at play in a donor's decisions in the field and successfully replicating them in a laboratory setting.

The discrepancy between expert predictions and the ground truth established in our field experiments (at least in this particular setting) also raises interesting questions (similar to those raised by DellaVigna and Pope (2017)) about the conditions under which expert intuitions will and will not predict field outcomes. A better understanding of how donors made their choices would enable a comparison with experts' theories of how donors make choices and may help answer these questions. Our evidence suggests that experts may have taken into account explicit trade-offs in motivation (e.g., that the threshold match had a lower match rate but incentivized donors to give more than before) but failed to consider opposing motivational forces that were implicit (e.g., the potentially detrimental effect of giving-credit framing on warm glow) or involved higher-order strategic behavior (e.g., the crowding-out effect in matching solicitations). Being more cognizant of these complex motivational factors might make experts better calibrated about their uncertainty when evaluating new fundraising appeals.

7.2. Implications for Fundraising Practices

Consider a fundraiser who relied on reasonable sources of information and chose to simply implement the giving-credit framing with threshold matching instead of running an experiment. What would this fundraiser learn from the experience? Unfortunately, the fundraiser would be unlikely to learn much from the observational data that result from just running a novel campaign

because of the lack of a comparison and the resulting inability to conclude reliably what the counterfactual donations would have been. As a result, fundraisers are unlikely to update their beliefs effectively, resulting in the perpetuation of the inaccurate expert opinions we have documented. Absent a controlled trial, as in our field experiment, the fundraiser would not have learned that the giving-credit intervention was a costly mistake.

The promise of theory-based decision making is that it provides generalizable guidance. However, many practitioners might not be surprised to hear that theoretical economic models developed by academics sometimes generate implications that do not hold up in the field. In fact, prior field experiments have documented other findings that may be rationalizable ex post under the impure altruism model with specific assumptions but that contradict plausible a priori interpretations of the model (Andreoni 1988, Eckel and Grossman 2003, Karlan and List 2007).

Practitioners who are skeptical of academic predictions may believe that they or their experienced colleagues will predict better. However, in our data, professional fundraising managers overwhelmingly and incorrectly predicted that the giving-credit framing would perform better than the regular framing. Other fundraising managers might put less trust in expert opinions but be confident that with a bit of data about donors, they could choose the best option. However, our low-cost incentive-compatible experiment with online participants, reported in the online appendix, provided directional evidence that the giving-credit framing was better than the standard framing. Thus, this kind of survey research also failed to identify the inferiority of the giving-credit framing.

Taken together, the results strongly support a pessimistic view of the fundraiser's ability to accurately predict actual outcomes without field experiment data in the relevant context. In this pessimistic view, we echo recent work that concludes that the combination of reasoning processes in a decision may result in different outcomes in different contexts (Goswami and Urminsky 2016a) and that even field results from one setting may not generalize to another (Alcott 2015). Although our findings about the weakness of the giving-credit framing relative to a typical match framing have been shown to be robust within a single context, replicated in two studies conducted three and half years apart, our data cannot speak to the robustness of the difference across different fundraising contexts. Thus, contrary to the conclusions of many academic papers, we do not recommend that fundraisers rely on our findings to make decisions about the type of framed match to use.

Instead, our positive recommendation for fundraisers is to simply test planned interventions in the field before full implementation and to continue testing

new ideas in the field, whether generated from theoretical models, empirical academic research, other fundraisers' seemingly successful practices, or expert intuitions. Concerns about unintended negative consequences of such experimentation, because of people's perceived dislike of experiments, seem quite unfounded (Mislavsky et al. 2019). Although such experiments are not costless, performing them is well within the reach of most fundraising organizations, and the learning can far outweigh the minimal costs (Goswami and Urminsky 2016b). In fact, many marketing organizations (particularly those operating online) have not only adopted experimentation but also have restructured their marketing activities around experimentation, to the point where field testing is an ongoing and seamless aspect of their everyday practice (Brynjolfsson and McAfee 2011). There is simply no substitute for in-context field experiments to test the consequences of fundraising interventions.

Acknowledgments

We thank the Science of Philanthropy Initiative for assistance with securing a field partner and the Hyde Park Art Center for conducting the field experiments. Research support from the UB Foundation and the University of Chicago Booth School of Business is gratefully acknowledged.

Endnotes

¹ Proportions do not add up to 100% because some people said "cannot predict."

² For example, see the 2018 Global Trends in Giving Report by Nonprofit Tech for Good, <http://rocklandcce.org/resources/2018-global-trends-in-giving-report>.

³ The raw distribution of contributions obtained in the various conditions is available in the online appendix.

⁴ For robustness, we also examined average contribution and net contribution using nonparametric tests. We also examined the raw responses after employing both Grubbs' test and winsorizing to handle outliers. The results are reported in the online appendix, as well as results after controlling for covariates.

⁵ The raw distribution of contributions obtained in the various conditions is available in the online appendix.

⁶ For robustness, we conducted nonparametric tests and used both a Grubbs' test and winsorizing to deal with outliers. The results are reported in the online appendix.

References

- Adena M, Huck S (2017) Matching donations without crowding out? Some theoretical considerations, a field, and a laboratory experiment. *J. Public Econom.* 148(April):32–42.
- Alcott H (2015) Site selection bias in program evaluation. *Quart. J. Econom.* 130(3):1117–1165.
- Andreoni J (1988) Privately provided public goods in a large economy: The limits of altruism. *J. Public Econom.* 35(1):57–73.
- Andreoni J (1990) Impure altruism and donations to public goods: A theory of warm-glow giving. *Econom. J.* 100(401):464–477.
- Andreoni J, Payne AA (2013) Charitable giving. Auerbach A, Chetty R, Feldstein M, Saez E, eds. *Handbook of Public Economics*, vol. 5, (Elsevier, Amsterdam), 1–50.

- Anik L, Norton MI, Ariely D (2014) Contingent match incentives increase donations. *J. Marketing Res.* 51(6):790–801.
- Baker RJ, Walker JM, Williams AW (2009) Matching contributions and the voluntary provision of a pure public good: Experimental evidence. *J. Econom. Behav. Organ.* 70(1):122–134.
- Bartels DM, Hastie R, Urminsky O (2018) Connecting laboratory and field research in judgment and decision making: Causality and the breadth of external validity. *J. Appl. Res. Memory Cognition* 7(1):11–15.
- Bartels DM, Kvaran T, Nichols S (2013) Selfless giving. *Cognition* 129(2):392–403.
- Bekkers R (2015) When and why matches are more effective subsidies than rebates. *Replication in Experimental Economics* (Emerald Group Publishing Limited, Bingley, UK), 183–211.
- Blumenthal M, Kalambokidis L, Turk A (2012) Subsidizing charitable contributions with a match instead of a deduction: What happens to donations and compliance? *Natl. Tax J.* 65(1):91–116.
- Crumpler H, Grossman PJ (2008) An experimental test of warm glow giving. *J. Public Econom.* 92(5–6):1011–1021.
- Davis DD, Millner EL, Reilly RJ (2005) Subsidy schemes and charitable contributions: a closer look. *Experiment. Econom.* 8(2):85–106.
- Dawes RM, Corrigan B (1974) Linear models in decision making. *Psych. Bull.* 81(2):95–106.
- Dawes RM, Faust D, Meehl PE (1989) Clinical vs. actuarial judgment. *Science* 243(4899):1668–1674.
- DellaVigna S, Pope D (2017) What motivates effort? Evidence and expert forecasts. *Rev. Econom. Stud.* 85(2):1029–1069.
- DellaVigna S, List JA, Malmendier U (2012) Testing for altruism and social pressure in charitable giving. *Quart. J. Econom.* 127(1):1–56.
- Dove KE (2000) *Conducting a Successful Capital Campaign: The New, Revised, and Expanded Edition of the Leading Guide to Planning and Implementing a Capital Campaign* (Jossey-Bass, San Francisco).
- Dubé JP, Luo X, Fang Z (2017) Self-signaling and prosocial behavior: A cause marketing mobile field experiment. *Marketing Sci.* 36(2):161–186.
- Eckel CC, Grossman PJ (2003) Rebate vs. matching: Does how we subsidize charitable contributions matter? *J. Public Econom.* 87(3):681–701.
- Eckel C, Grossman PJ (2017) Comparing rebate and matching subsidies controlling for donors' awareness: Evidence from the field. *J. Behav. Experiment. Econom.* 66(February):88–95.
- Escalas JE, Bettman JR (2003) You are what they eat: The influence of reference groups on consumers' connections to brands. *J. Consumer Psych.* 13(3):339–348.
- Feiler DC, Tost LP, Grant AM (2012) Mixed reasons, missed givings: The costs of blending egoistic and altruistic reasons in donation requests. *J. Experiment. Soc. Psych.* 48(6):1322–1328.
- Fitzsimons GJ, Lehmann DR (2004) Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Sci.* 23(1):82–94.
- Frey BS, Meier S (2004) Social comparisons and pro-social behavior: Testing “conditional cooperation” in a field experiment. *Amer. Econom. Rev.* 94(5):1717–1722.
- Goswami I, Urminsky O (2016a) When should the ask be a nudge? The effect of default amounts on charitable donations. *J. Marketing Res.* 53(5):829–846.
- Goswami I, Urminsky O (2016b) When and how fundraisers should suggest how much to donate. *Harvard Bus. Rev.* (December 29), <https://hbr.org/2016/12/when-and-how-fundraisers-should-suggest-how-much-to-donate>.
- Harbaugh WT (1998) The prestige motive for making charitable transfers. *Amer. Econom. Rev.* 88(2):277–282.
- Harrison GW, List JA (2004) Field experiments. *J. Econom. Lit.* 42(4):1009–1055.
- Huck S, Rasul I (2011) Matched fundraising: Evidence from a natural field experiment. *J. Public Econom.* 95(5):351–362.
- Huck S, Rasul I, Shephard A (2013) Comparing charitable fundraising schemes: Evidence from a natural field experiment and a structural model. *American Econom. J. Econom. Policy* 7(2):326–369.
- Karlan D, List JA (2006) *Does price matter in charitable giving?* Evidence from a large-scale natural field experiment. NBER Working Paper No. 12338, National Bureau of Economic Research, Cambridge, MA.
- Karlan D, List JA (2007) Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *Amer. Econom. Rev.* 97(5):1774–1793.
- Karlan D, List JA (2012) *How can Bill and Melinda Gates increase other people's donations to fund public goods?* NBER Working Paper No. 17954, National Bureau of Economic Research, Cambridge, MA.
- Karlan D, List JA, Shafir E (2011) Small matches and charitable giving: Evidence from a natural field experiment. *J. Public Econom.* 95(5):344–350.
- Landry CE, Lange A, List JA, Price MK, Rupp NG (2006) Toward an understanding of the economics of charity: Evidence from a field experiment. *Quart. J. Econom.* 121(2):747–782.
- Lange A, List JA, Price MK (2007) Using lotteries to finance public goods: Theory and experimental evidence. *Internat. Econom. Rev.* 48(3):901–927.
- Latané B, Williams K, Harkins S (1979) Many hands make light the work: The causes and consequences of social loafing. *J. Personality Soc. Psych.* 37(6):822–832.
- Leonhardt D (2008) What makes people give? *New York Times* (March 9), <https://www.nytimes.com/2008/03/09/magazine/09Psychology-t.html>.
- List JA, Lucking-Reiley D (2002) The effects of seed money and re-funds on charitable giving: Experimental evidence from a university capital campaign. *J. Political Econom.* 110(1):215–233.
- List JA, Price MK (2009) The role of social connections in charitable fundraising: Evidence from a natural field experiment. *J. Econom. Behav. Organ.* 69(2):160–169.
- List JA, Murphy JJ, Price MK, James AG (2019) Do appeals to donor benefits raise more money than appeals to recipient benefits? Evidence from a natural field experiment with Pick.Click.Give. NBER Working Paper No. 26559, National Bureau of Economic Research, Cambridge, MA.
- Martin R, Randal J (2009) How Sunday, price, and social norms influence donation behaviour. *J. Socio-Econom.* 38(5):722–727.
- Meier S (2007) Do subsidies increase charitable giving in the long run? Matching donations in a field experiment. *J. Eur. Econom. Assoc.* 5(6):1203–1222.
- Meier S, Frey BS (2004) Matching donations: Subsidizing charitable giving in a field experiment. Zurich IIEER Working Paper No. 181.
- Mislavsky R, Dietvorst BJ, Simonsohn U (2019) The minimum mean paradox: A mechanical explanation for apparent experiment aversion. *Proc. Natl. Acad. Sci. USA* 116(48):23883–23884.
- Oppenheimer D, Meyvis T, Davidenko N (2009) Instructional manipulation checks: Detecting satisficing to increase statistical power. *J. Experiment. Soc. Psych.* 45(4):867–872.
- Ribar DC, Wilhelm MO (2002) Altruistic and joy-of-giving motivations in charitable behavior. *J. Political Econom.* 110(2):425–457.
- Rondeau D, List JA (2008) Matching and challenge gifts to charity: Evidence from laboratory and natural field experiments. *Experiment. Econom.* 11(3):253–267.
- Sanders M, Smith S, Norton MI (2013) Non-standard matches and charitable giving other. Harvard Business School Working Paper, No. 13–094, May 2013.
- Shang J, Croson R (2009) A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *Econom. J.* 119(540):1422–1439.
- Shen L, Hsee CK, Wu Q, Tsai CI (2012) Overpredicting and underpredicting in pricing decisions. *J. Behav. Decision Making* 25(5):512–521.
- Tonin M, Vlassopoulos M (2010) Disentangling the sources of pro-socially motivated effort: A field experiment. *J. Public Econom.* 94(11–12):1086–1092.
- Zhang CY, Sussman AB (2018) Perspectives on mental accounting: An exploration of budgeting and investing. *Financial Planning Rev.* 1(1–2):e1011.