Oxford Economic Papers, 2023, **00**, 1–19
https://doi.org/10.1093/oep/gpad021
Article

OXFORD

What drives overhead aversion in charity? Evidence from field-experimental variation in fundraising costs

Tobias Cagala[1], Johannes Rincke[2,*], Amanda Tuset Cueva[2]

[1]Deutsche Bundesbank, Frankfurt am Main 60006, Germany
[2]Department of Economics, University of Erlangen-Nuremberg, Nuremberg 90403, Germany
*Corresponding author: Email: johannes.rincke@fau.de

Abstract

This article explores donors' aversion to financing charities' fundraising expenses. We hypothesize that such expenses can signal a charity's efficiency, or affect the donors' perception of the impact a donation has on the cause. Using data from a randomized field experiment, we disentangle both effects, differentiating between weakly and strongly committed donors. Among potential donors who are weakly committed to the cause, information on the charity's efficiency does not affect donation behavior. Signaling an increased impact leaves unaffected the average donation among weakly committed donors, but diminishes their likelihood to give. Regarding strongly committed donors, we find that impact-related information does not affect behavior, but a signal of improved efficiency strongly increases donations along the intensive margin. We conclude that information on fundraising expenses plays little role for weakly committed donors. In contrast, strongly committed donors are averse to financing fundraising expenses mostly due to efficiency concerns.

JEL classifications: D64, C93

1. Introduction

The 25 largest US charities spend between 5% and 25% of total donations on fundraising expenses (Andreoni and Payne, 2011). One reason why high fundraising costs are a matter of concern for charities is an aversion among donors to finance overhead (Tinkelman and Mankaney, 2007; Caviola *et al.*, 2014).[1] Whereas the sensitivity of donors with respect to overhead is well documented (Callen, 1994; Tinkelman and Mankaney, 2007; Trussell and Parsons, 2007; Caviola *et al.*, 2014), the channels through which overhead-related information affects donors' behavior are not fully understood. Two potential channels have been discussed (Gneezy *et al.*, 2014). First, donors might exploit overhead-related information for inference on the charity's efficiency, or quality more broadly. The motive for doing so would be an aversion against the wasteful spending of donated funds. Second, overhead might affect the donors' perceptions of a donation's impact on the cause. Here, the motive would be to personally make a difference.

[1] It has long been acknowledged that donors should not, in theory, evaluate charities according to their overhead ratios, but rather focus on charities' cost–effectiveness.

This article exploits a randomized field experiment to explore the relevance of both channels and to contribute to a better understanding of overhead aversion in charity. We partnered with the Protestant Church in Bavaria, Germany, and two of its local church districts. We focus on fundraising expenses as a subset of all overhead. In the experiment, the church districts communicated a change in fundraising costs in solicitation letters that ask church members to donate to a district-specific local church fund. Our treatments frame a given refund scheme for fundraising costs borne by the districts in different ways and thereby allow us to separately identify both the efficiency and the impact channel. First, we measure the church members' response to a signal of improved efficiency while keeping constant the impact a donation has on the cause. Second, we identify the effect of increasing the impact of a donation while keeping constant the charity's efficiency. The church sent out solicitations to all adult church members, irrespective of their donations in previous years. This allows us to trace out the heterogeneity of the treatment responses between potential donors who, by their baseline behavior, have revealed their degree of commitment to the cause. Given that we rely on treatments that frame a given refund scheme differently, a caveat is that our interpretations rest on assumptions about how the church members perceived the different letters. We discuss the assumptions and possible concerns about deception in Section 3.

In part, our results consider the heterogeneity in treatment responses regarding the degree of individuals' commitment to the cause. For that purpose, we exploit the fact that the solicitation letters used by the church display an income-dependent schedule of suggested donation amounts. Using data on donations in the years prior to the intervention that we can link to income data, we predict for each potential donor whether the individual would have donated the personalized suggested amount (or more) without an intervention. We label potential donors who are predicted to follow the suggestion as *strongly committed* donors, and individuals predicted not to follow the suggestion as *weakly committed* donors. Our definition of donor types is a refinement of the common distinction between individuals who donated in the past to the given cause ('warm list') and individuals who did not ('cold list').

Considering as outcomes an indicator for donors who give more than the amount suggested in the solicitation letter (i.e. particularly generous donors), a donation indicator, and the donation amount, our analysis delivers two sets of main results. Our first main finding relates to weakly committed potential donors. We show that by signaling a reduction in fundraising expenses, the fundraiser is unable to positively affect the donation behavior of this donor type. Whereas signaling that the reduction in fundraising expenses has improved the charity's efficiency does not affect donor behavior across the three outcomes, communicating an increased impact of donations on the cause actually crowds out the likelihood to give among weakly committed donors. The latter finding may be counterintuitive, but is consistent with economic theory provided that weakly committed donors are motivated by altruism. This is because a lowering of fundraising expenses is equivalent to a third-party transfer to the public good provided by the charity. Hence, our finding on weakly committed donors is in line with the public-goods crowding-out hypothesis (Warr, 1982; Roberts, 1984; Steinberg, 1986; Andreoni, 1989), which states that donors with altruistic preferences will reduce their private donations in response to a grant from a third party to the public good.

Our second main finding is that donors who are strongly committed to the cause ignore impact-related information (no significant effects across all three outcomes), but respond positively to a signal of improved efficiency along the intensive margin. The positive efficiency effect on giving among strongly committed donors is economically sizable: Relative to their counterparts in the control group, strongly committed donors give 21.1% more on average, and their likelihood to donate more than the suggested amount increases from 28.5% to 42.7%. In contrast, the likelihood to donate among strongly committed donors is not affected by the efficiency treatment.

The positive impact of the efficiency treatment on the likelihood to donate more than the suggested amount among strongly committed donors is strong enough to significantly shift this outcome even if we consider all donors in our sample jointly. In contrast, neither the likelihood to donate nor the donation amount is significantly affected on average by the efficiency treatment. Similarly, using a reduction in fundraising expenses to increase the impact a donation has on the cause does not affect any of the outcomes on average. Obviously, the latter findings depend not only on the size of the treatment effects for the two donor types but also on the share of strongly committed donors in the sample. As the latter parameter will differ a lot between contexts, we do not want to put too much emphasis on the average treatment effects.

The literature has widely documented overhead aversion in charity (Callen, 1994; Tinkelman and Mankaney, 2007; Trussell and Parsons, 2007; Caviola *et al.*, 2014), but produced little evidence on why changes in overhead costs affect the behavior of real-world donors.[2] Most closely related to our work is Gneezy *et al.* (2014), who run a laboratory and a field experiment to shed light on donors' overhead aversion. In the laboratory experiment, subjects decide which of two charities should receive a $100 donation, and the treatments aim at disentangling the efficiency and impact motives. The authors reject the efficiency motivation and find support for the impact motivation. The field experiment uses treatments that vary the information on how a charity uses a given amount of seed money. The authors find that using seed money to fully cover a project's fundraising costs raises more money relative to using it to finance a matching scheme or just communicating the existence of a lead donation. The field experiment demonstrates that overhead aversion is a relevant motive, but the design does not allow to disentangle the channels through which overhead aversion works. In contrast, our key contribution is evidence on the relative importance of the impact and the efficiency motive. Unlike Gneezy *et al.*'s (2014) laboratory experiment, we find no evidence that overhead aversion works through the impact channel. In our setting, strongly committed donors are mainly motivated by efficiency concerns, and we further identify a crowding-out effect of the impact treatment in the sample of weakly committed donors. One possible explanation for these differences is contextual factors, which naturally vary strongly between the laboratory and the field. For instance, the fact that the donors in our setting have a long-term relationship with the church as a charitable organization may affect how donors perceive fundraising costs relative to a one-time decision to donate in the laboratory.

Other previous work on the role of overhead in charity includes, for example, Okten and Weisbrod (2000), Bowman (2006), and Meer (2014). These studies use administrative data on cross sections of charities to explore how donors respond to changes in the price of giving that are induced by variations in overhead.

As pointed out by, for example, Vesterlund (2003) and Bekkers and Wiepking (2011), the evidence on leadership donations, seed money, and matching schemes (for example, List and Lucking-Reiley, 2002; Karlan and List, 2007; Landry *et al.*, 2010; Huck and Rasul, 2011; Huck *et al.*, 2015) likely reflects that donors use the observable behavior of others to draw inferences about the quality of charities. Our findings allow for a similar interpretation at least for strongly committed donors: Donors of this type may mainly use overhead-related information to update their beliefs about the likelihood that the charity will deliver on its promises. Regarding strongly committed donors, our work thus supports the view that a major force shaping donation behavior is a preference to give to high-quality charities.

[2] There is substantial variation in how the literature has framed overhead aversion and in empirical measurement. Callen (1994) addresses charity efficiency, relating measures for charity output to inputs of volunteer hours. Similarly, Trussell and Parsons (2007) and Tinkelman and Mankaney (2007) exploit financial reporting figures to measure efficiency. Caviola *et al.* (2014) use overhead ratio and cost-effectiveness to study how donors evaluate competing charities.

The remainder of the article is organized as follows. Section 2 presents the setting of the field experiment, Section 3 describes the experimental design and the data, and Section 4 discusses our findings. Section 5 concludes.

2. Setting of the field experiment

To implement our research design, we collaborated with the Protestant Church in Bavaria, Germany, and two of its local administrative units, the church districts. In total, the two districts comprise 14 urban parishes with about 35,000 adult church members. In a coordinated fundraising drive, each Bavarian church district sends out a solicitation letter once a year to all individual resident church members above the age of 18 years. The letter asks church members to contribute to a district-specific local church fund. The annual fundraising drive is in addition to the collection of the state church tax church members are liable for (see the following paragraphs for details). The solicitation letters, furthermore, contain a leaflet that provides examples of recent and ongoing projects funded by donations. These are mostly projects related to renovations and repairs of church buildings, like the refurbishment of historic church windows and organs.

A distinctive feature of our setting is that the solicitation letter contains an income-dependent schedule of suggested donation amounts, displayed in Table 1. The schedule lists suggested amounts ranging from €5 to €100 for a total of six income brackets. This schedule emerges from the fact that the German state allows the churches to raise so-called 'church taxes' from their members.[3] Referring to this provision, the solicitation letter asks the recipients to use the schedule for a self-assessment of income, and to transfer the suggested amount corresponding to the recipient's income bracket to the fund's bank account.[4] Importantly, because of the income-dependent schedule, the suggested donation amount varies across letter recipients. Moreover, determining an individual donor's suggested amount requires information on the individual's income. The church districts have no information on individual incomes. As a result, the suggested amount is the donor's private information, and the districts have no means to enforce that members pay the suggested amounts (or pay anything, for that matter).

Previous research has documented that the absence of enforcement regarding the church contribution is well-known among church members (Dwenger *et al.*, 2016). Also, the church districts do not share information on individual payments with priests or staff working for the parishes belonging to a given church district. As a consequence, whether or not church members contribute to the local church fund, and (if they make a payment) whether they pay the suggested amount or not, does neither affect access to church services nor the members' social standing with their local congregation. As a result, the setting of our experiment is similar to other charitable-giving contexts where potential donors respond to solicitation letters, and individual giving is private information. This is also reflected in average payment behavior absent any treatment: In the no-intervention group, only 19.8% of church members donate to the local church fund in the treatment year in response to the solicitation.[5] While we acknowledge that our context is specific and deviates from other pure

[3] The Protestant Church in Bavaria finances itself mainly through a state church tax which corresponds to 8% of church members' income tax liabilities. Tax collection is organized at the state level, and the state church uses a grant system to channel part of the tax revenue to the church districts and parishes. The church raises about €270 in state church taxes per church member (including members not paying any church taxes) and year. Church members can avoid paying the state church tax by opting out of their church membership at any time. The church tax is therefore akin to a recurring donor scheme (Bittschi *et al.*, 2021).

[4] The wording is as follows: 'The local church contribution is staggered according to income. Please self-assess your income using the adjoining schedule and transfer your contribution within a month's time'.

[5] When we pre-registered the experiment, we referred to the context as a tax setting with zero enforcement and highlighted the role of tax morale. Responding to the referee's comments that payments in the given institutional setting are more properly described as donations, we changed the paper's frame to charitable giving. We believe that studying the given setting has relevance for both fields.

Table 1. Schedule of personalized suggested donation amounts.

Annual income	Annual suggested amount	Percentage of sample in income bracket (Conditional on income being observed)
€8,005–9,999	€5	2.9
€10,000–24,999	€10	37.4
€25,000–39,999	€25	34.4
€40,000–54,999	€45/€50	13.5
€55,000–69,999	€70	5.6
€70,000 and above	€100	6.3

Notes: This table shows the schedule of the suggested donation amount in the two districts where the field experiment was implemented. In one of the districts, potential donors falling into the bracket between €40,000 and €54,999 face a suggested amount of €45, while in the other district the respective suggested donation amount is €50.

charitable-giving contexts in many important respects, the rate of church members donating at baseline seems well in line with other settings where potential donors first self-select into a donor base (which would be church membership in our case), and then (conditional on self-selection) make a donation decision in response to recurring fundraising drives among individuals in the base.[6] It is also worth noting that in Germany (allegedly), wasteful spending by the (Protestant and Catholic) churches is a recurring theme within churches and beyond. In response to demands for better transparency and accountability in financial matters, the Protestant Church in Bavaria set up a website presenting information on the topic 'Money and the Church'.[7]

While the church districts have no information on individual incomes, the church administration at the state level receives individual income records from the state's tax authorities on all church members who file for the federal income tax. These income records are used to determine the church members' church taxes, the main source of church revenue. An advantage of our setting is that the church allowed us to access the income data and link them to individual donations to the local church funds.[8] This means that, for the sample of church members who file for the federal income tax, for the purposes of the research project, we could determine the suggested donation amount implied by the donation schedule highlighted in the solicitation letter, and compare the suggested amount to the actual donation. Hence, despite the fact that the suggested donation amount is differentiated by income and unknown to the church districts, in the linked data, we do observe whether individuals follow the suggestion or not. This allows us to control very flexibly for income effects in our estimations, and to construct measures describing how strongly committed to the cause individuals are (see next section for details).

3. Experimental design, data, and methods

In the following, we discuss the field-experimental setup. We first describe the hypotheses and the experimental design, then discuss the data and sampling, and finally comment on how we distinguish between strongly and weakly committed donors.

[6] Other examples of self-selection of potential donors into a donor base include, for instance, university alumni associations. For further discussion, see Bittschi *et al.* (2021).

[7] See https://www.kirche-und-geld.de (website only in German).

[8] The procedures used to link both data sources were reviewed and approved by state church officials. In addition, the experimental procedures and the handling of the data were approved by the Ethics Committee of the School of Business and Economics at the University of Erlangen-Nuremberg. All data were treated strictly confidential, and no information on individual incomes was shared with the districts or any other third party.

3.1 Experimental design

3.1.1 Hypotheses

The aim of the experiment is to explore why donors are averse to financing fundraising expenses. Following Gneezy *et al.* (2014), we hypothesize that donors respond to information on fundraising expenses because they perceive the information as a signal either about the charity's efficiency or about the impact a donation has on the cause. Of course, donors might be motivated by concerns regarding efficiency and impact at the same time. Our treatments are meant to disentangle both effects empirically and determine their relative importance for the behavior of donors. Our first hypothesis refers to the efficiency effect:

> H1: The higher a charity's efficiency, the more willing potential donors are to give to that charity, holding constant the impact a donation has on the cause.

In contrast, our second hypothesis relates to the effect a donation has on the cause:

> H2: The higher a donation's impact on the cause, the more willing potential donors are to give to that charity, holding constant the charity's efficiency.

Hypothesis H2 describes a behavioral response that is due to a possible preference of donors to give to high-impact charities (i.e. donors are more willing to give if their donation 'makes a difference'). As in Gneezy *et al.*'s (2014) field experiment, our experiment considers a change in impact that is achieved by an infusion of external funds used to cover fundraising expenses. A large literature initiated, for example, by Warr (1982), Roberts (1984), and Andreoni (1989) has shown theoretically that, if preferences are altruistic and a charity receives a grant from a third party, this grant fully crowds out private donations. According to this theory, an increase in the impact a donation has on the cause that is achieved by seed money should not crowd-in, but crowd-out donations. If we find that the impact effect is negative (crowding-out of donations), this would not only suggest the absence of a behavioral motivation to make a difference but also confirm one of the key implications of the public-goods crowding-out hypothesis.

3.1.2 Implementing changes in fundraising expenses

The solicitation letters were mailed by the church district administrations to all resident church members in June 2016. All letters were mailed on the same day. In terms of layout and general content, all letters (see the Supplementary Appendix for a sample) were similar to the letters mailed in previous years. The letter highlights that the suggested donation amount depends on the church member's income and asks the recipient to determine the appropriate amount using the schedule of six income brackets.

The experimental conditions varied only in a short note referring the recipients to a change in the charity's expenses associated with the fundraising drive. Before turning to the details of the experimental conditions, we would like to highlight that the changes in fundraising expenses communicated in all letters were achieved through a scheme that refunded the church districts participating in the experiment for part of the costs associated with the mailing of the letters.

The details of the refund scheme were as follows. In preparing the solicitation letters, the districts rely on in-house service providers operating within the state-level administration of the church. The service providers handle the records of church members, produce mailings on behalf of the district administrations, and deliver them to a mail service company. The service providers bill the districts for their services. In the experiment, those bills represent the fundraising costs borne by the districts. In collaboration with the different tiers of the church administration, we arranged for a partial reimbursement of the districts for the billed cost through state-level church funds. For each district, the overall reimbursement

reflected exactly the distribution of treatments and the changes in locally borne fundraising costs communicated in the respective letters (see next paragraph for details). As a result, each individual letter carried a change in fundraising costs borne by the district in accordance with the respective treatment paragraph.

The fundraising we study is relevant for the districts because they keep 100% of the net revenues. Specifically, the revenues from the annual fundraising drive are exempt from the fiscal equalization scheme that the state church uses to determine the districts' and the associated parishes' budgets. This means that net revenue changes from the fundraising we study are not compensated by the state church in any way. This was also true in the experiment: The changes in fundraising expenses (i.e. the cost refunds) were fully effective for the local church districts, with no compensation of any form in some other layer of the financial relations between the state church and local church districts.

3.1.3 Letters

In the following, we discuss the experimental conditions that allow us to separately identify the efficiency and the impact effect.

To identify the efficiency effect, we compare donations in an efficiency treatment to a control group. As a signal of improved efficiency, the letter in the respective treatment states the following: 'This year, we were able to reduce the administrative cost associated with the mailing of this letter by 30%. This means that your contribution is now even more effective'. As described before, the districts' fundraising expenses were in fact reduced in accordance with this statement. However, the letter in the efficiency treatment did not point the recipients to the fact that the reduction in fundraising expenses was made possible by a refund from the state church. For the comparison to identify the efficiency effect, in the control group, we need to hold constant (relative to the efficiency treatment) the impact of a donation. Hence, the control group letter also needs to communicate a 30% change in fundraising costs, but in a way that does not signal anything about the efficiency of the church as a charitable organization. To achieve this goal, the control group letter points the recipient to the refund scheme in place and thus frames the change as shifting of fundraising costs from the district to the state church: 'This year, we get a refund from the state church that covers 30% of the administrative cost associated with the mailing of this letter. This means that your contribution is now even more effective'. To summarize, the efficiency effect is identified by framing an existing refund scheme in different ways.

Given that the efficiency effect is identified by framing a given manipulation of fundraising expenses in different ways, the question arises whether the efficiency treatment was truthful. In the following, we discuss possible concerns about potential deception, recognizing that the boundaries of deception are not set in stone. A first point to mention is that, as discussed before, the changes in fundraising expenses communicated in the letters all became fully effective. Hence, a given donation allowed the church to supply more of the public good. Second, regarding the means by which the reductions in fundraising expenses were achieved, there is no question that the control group letter (which stated that the reduction in fundraising expenses was implemented via a refund scheme) was truthful. As regards the letter in the efficiency treatment, the refund scheme was not mentioned. A possible concern could be that as a result, (some) recipients may have perceived the letter in the efficiency treatment as suggesting a permanent (rather than a one-time) reduction in fundraising costs. Given that the refund scheme was only implemented in the year when the experiment took place, this could be judged as being deceptive. We acknowledge that we cannot know how the subjects perceived the treatments, and we cannot preclude that some subjects (falsely) believed the change in fundraising expenses to be permanent. We would like to note that, if anything, the phrasing of the treatment paragraph ('This year, we were able…') pointed the recipients to the one-time character of the change in fundraising expenses. Moreover, the most obvious way how the church could actually achieve the cost reduction communicated

in the treatment would be to contract with service providers outside the church administration, instead of using the in-house providers. Whereas the in-house providers charge the districts a piece rate of €0.72 per solicitation letter mailed, the market price is closer to about €0.50, implying a possible cost reduction of about 30%. Hence, a district that would switch from in-house to external provision could achieve the reduction in fundraising expenses communicated in the efficiency letter. Of course, the decision to contract with an external provider could be revoked at any time, or providers could increase prices in later years. As a result, switching to an external provider would *not* necessarily imply a permanent cost reduction.

To reiterate, we acknowledge that we do not know how exactly the church members perceived the letters. Despite the fact that the statements in the different treatment paragraphs were technically true, it is still possible that participants made incorrect inferences about what the efficiency message implied. As a result, we rely on assumptions when interpreting our results. The key assumptions are, first, that the recipients in the efficiency treatment did perceive the letter as communicating an improved efficiency of fundraising as a result of activities of the local church administration directed toward that goal. Second, we need to assume that the control group letter was not perceived as a signal about the efficiency of the church as a charitable organization.

To separately identify the effect of changing the impact of a donation on the cause, we employ a treatment that communicates a 100% shift of fundraising costs to the state church and compare it to the control group with its 30% cost shift. The letter in the impact treatment states the following: 'This year, we get a refund from the state church that covers all administrative costs associated with the mailing of this letter. This means that your contribution is devoted to its purpose without any deduction. Hence, your contribution is now even more effective'. Importantly, between the impact treatment and the control group, the source of the change in fundraising costs is kept constant. Hence, the donors' perception of the efficiency of the church as a charitable organization should not be differentially affected. We would like to highlight that the transfer covering the fundraising cost in the impact treatment can be interpreted as seed money that is used to cover the overhead of a fundraising drive. The impact treatment is thus very similar to the overhead treatment in Gneezy *et al.*'s (2014) field experiment.[9] While Gneezy *et al.* (2014) compare their overhead treatment to alternative uses of seed money, we identify the impact effect by varying the fraction of fundraising costs covered by a third party.

Finally, following the same randomization scheme, we also sampled potential donors into a no-intervention group. The respective letter omitted the paragraph on fundraising costs and, thus, was very similar to the letters sent in previous years. The purpose of the no-intervention group is to train a model that predicts individual donation behavior in the treatment and control groups absent any treatment (see Section 3.3 for details). In our empirical analysis, the predicted donor type is used to test for heterogeneous treatment effects.[10]

3.2 Data and sampling

3.2.1 Data sources

The church administration provided us with individual characteristics on all adult church members residing in either of the two districts. We also obtained individual-level records of

[9] The overhead treatment in Gneezy *et al.* (2014) states that 'a private donor who believes in the importance of the project has given this campaign a grant in the amount of $10,000 to cover all the overhead costs associated with raising the needed donations'.

[10] We do not compare outcomes between the no-intervention group and the other groups. As outlined before, the letter in the no-intervention group is used to predict donor types in the remaining groups and does not mention the charity's fundraising expenses. Since the letters in control, efficiency, and impact all refer the recipient to fundraising expenses, whereas the no-intervention letter does not, the comparison between any of the three letters used to identify the treatment effects and the no-intervention letter would vary two aspects at the same time: the specific information on reductions in fundraising expenses, and the salience of these expenses. The comparisons would thus be confounded by differences in the salience of fundraising costs.

donations for the 2 years preceding the treatment (baseline) and the treatment year. While the data for the baseline years are used to predict donation behavior absent any treatment, the data for the treatment year is used to derive our outcomes of interest. The third data source consists of individual-level income records, normally used by the state church's tax office to determine the church members' church tax. These data allow us to infer the personalized suggested donation amount for each potential donor, provided that income is observed. In our analysis, we can thus compare personalized suggested donation amounts to actual donations.

3.2.2 Sampling

The solicitation letters were sent to all resident adult church members. The experimental sample excludes individuals living in households with more than one adult church member, because household members often combine their individual donations into a single bank transfer, introducing measurement error in outcomes. In additional analyses reported in Section 4.3, we demonstrate that this decision is inconsequential to our findings. We also excluded employees of the Protestant Church from the experiment, since we cannot preclude that they learned about the experiment before implementation through their professional network. Using a stratified randomization scheme, we assigned each church member to one of the four letter groups.[11]

We focus on the sample of church members for whom we observe the personalized suggested donation amount (i.e. individual income) in the baseline.[12] This sample comprises 8,617 potential donors in total, out of which 6,433 individuals belong to the estimation sample comprising the efficiency treatment, the impact treatment, and the control group (the remaining 2,184 individuals belong to the no-intervention group). Note that we could not fully determine this sample ex ante, as income tax declarations in Germany are often filed with a considerable time lag. Hence, we could not determine ex ante the sample of church members for whom baseline income information would be available *ex post*. We, therefore, sampled all church members (subject to the restrictions discussed before) and restricted the sample *ex post* to those potential donors with sufficient income information to determine the personalized suggested donation amount in the baseline. A second restriction applies if we condition the sample on income information being available not only in the baseline but also in the treatment year. The part of the estimation sample which can be used for the respective regressions comprises 3,625 individuals.[13] Table 2 shows descriptives on baseline characteristics and corresponding balancing checks for the sample of potential donors whose type can be determined. Table A1 in the Supplementary Appendix repeats the balancing checks on the sample of subjects with a complete set of outcomes.

A possible concern regarding external validity could be that church members might differ from other potential donors in important respects. For instance, it is often claimed that Protestants have a special work ethic, which could translate into higher incomes and specific attitudes toward (some forms of) charitable causes. It could also be that church members are more pro-social in general. Table A2 in the Supplementary Appendix sheds light on these concerns and shows that Protestants are broadly similar to the overall population in Germany, and also similar to non-church members. If anything, members of the Protestant Church earn slightly lower incomes, most likely reflecting a selection effect resulting from a

[11] Strata were defined by age, gender, income bracket (including a dummy for taxpayers for whom we did not observe the true tax liability at the time of sampling), baseline donation behavior, and parish.

[12] This essentially limits our sample to church members who file an income tax declaration. For non-filers (comprising most students and retirees), no income data is transmitted to the church tax office. We also lose observations from imperfections in the match between donations and income records.

[13] There are two reasons why we observe income for fewer individuals in the treatment year: First, we can predict a donor's type even if income is only observed in one of the baseline years (see the next subsection for details). Second, we had to collect all income data in the year following the intervention, implying that late filing leads to more missing values in the income records for the treatment year relative to the baseline years.

Table 2. Baseline characteristics.

	T0: Control	T1: Efficiency	T2: Impact	Difference T1−T0	Difference T2−T0
A: Full sample					
Female	0.496	0.493	0.487	−0.003	−0.010
				(0.015)	(0.015)
Age	48.8	48.6	48.8	−0.154	0.012
				(0.494)	(0.493)
Suggested amount ≤ €10	0.469	0.460	0.473	−0.009	0.004
				(0.015)	(0.015)
€25 ≤ Suggested amount ≤ €50	0.411	0.421	0.407	0.009	−0.005
				(0.015)	(0.015)
Suggested amount ≥ €70	0.120	0.120	0.120	0.000	0.001
				(0.010)	(0.010)
Donated amount	7.95	8.28	8.28	0.332	0.337
				(0.643)	(0.636)
Makes donation	0.225	0.218	0.222	−0.007	−0.003
				(0.013)	(0.013)
Number of observations	2156	2130	2147	4286	4303
B: Weakly committed					
Female	0.481	0.478	0.477	−0.004	−0.004
				(0.017)	(0.017)
Age	46.5	46.4	46.5	−0.065	0.062
				(0.488)	(0.489)
Suggested amount ≤ €10	0.455	0.450	0.464	−0.005	0.009
				(0.016)	(0.017)
€25 ≤ Suggested amount ≤ €50	0.422	0.430	0.413	0.008	−0.009
				(0.016)	(0.016)
Suggested amount ≥ €70	0.123	0.120	0.123	−0.003	−0.001
				(0.011)	(0.011)
Donated amount	3.05	3.07	2.99	0.029	−0.054
				(0.423)	(0.446)
Makes donation	0.095	0.096	0.084	0.001	−0.011
				(0.010)	(0.009)
Number of observations	1,839	1,828	1,808	3,667	3,647
C: Strongly committed					
Female	0.584	0.589	0.540	0.006	−0.044
				(0.040)	(0.039)
Age	62.2	62.0	60.8	−0.111	−1.335
				(1.410)	(1.386)
Suggested amount ≤ €10	0.552	0.520	0.522	−0.032	−0.030
				(0.040)	(0.039)
€25 ≤ Suggested amount ≤ €50	0.350	0.364	0.372	0.014	0.022
				(0.039)	(0.038)
Suggested amount ≥ €70	0.098	0.116	0.106	0.018	0.008
				(0.025)	(0.024)
Donated amount	36.38	39.77	36.50	3.40	0.13
				(2.61)	(2.35)
Makes donation	0.978	0.957	0.956	−0.021	−0.022
				(0.014)	(0.014)
Number of observations	317	302	339	619	656

Notes: The table displays means of potential donors' baseline characteristics, together with estimated differences in means and corresponding standard errors. The sample consists of all potential donors for whom baseline income is observed. *Source*: Authors' calculations.

higher likelihood among high-income church members to opt out of their membership and thereby avoid the church tax (Bittschi *et al.*, 2021). The average amount of all charitable donations made in a given tax year is very similar across the different groups, suggesting that Protestants do not, on average, differ from the German population in terms of their general willingness to give to charities. As a caveat, one should keep in mind that our sample consists of individuals living in households with just one adult church member. Hence, we study a somewhat specific population of church members.[14]

3.3 Predicting donor types

In the charitable giving literature, it is common to distinguish between warm-list and cold-list individuals. Much of the recent experimental work on individual donation decisions covers only warm-list individuals, and when broader samples are considered, it is common to discuss the respective subsamples separately (Landry *et al.*, 2010). The feature of a personalized suggested donation amount, and the fact that we can observe whether donors follow the suggestion or not, allows us to refine the traditional warm-list versus cold-list distinction. Specifically, we exploit the available information to single out potential donors who are strongly committed to the cause (as opposed to weakly committed donors), and then explore the heterogeneity in treatment responses by the degree of the donors' commitment.

Our approach to identifying strongly committed donors is to predict for each potential donor whether the individual would have donated the personalized suggested amount (or more) without an intervention. Hence, our criterion when defining 'strong commitment to the cause' is whether the potential donor is predicted to follow the personalized suggestion formulated in the solicitation letter. To make this prediction, we use the observations in the no-intervention group to train a parsimonious model linking the donor's type (strongly versus weakly committed) in the year of the intervention to information on the donor's characteristics and her type (defined as an indicator for subjects who donated the personalized suggested amount, or more) in the two baseline years. After estimating the model, we predict each donor's type (out-of-sample) in the control group and the treatment groups. The predicted type is used to distinguish between strongly and weakly committed donors.

The probit regression[15] used to predict individual i's type (1 for strongly committed donor, 0 otherwise) in the treatment year t reads

$$Prob(TYPE_{i,t}|\cdot) = Prob(\delta_0 + \delta_1 TYPE_{i,t-1} + \delta_2 TYPE_{i,t-2} + X_{i,t}'\pi > u_{i,t}), \tag{1}$$

where $TYPE_{t-1}$ and $TYPE_{t-2}$ are indicators for the type in the baseline years $t-1$ and $t-2$, respectively, and X_t denotes the vector of strata variables.

The regression results (see Table A3 in the Supplementary Appendix) show that $TYPE_{t-1}$ and $TYPE_{t-2}$ are strong predictors of the type in the treatment year, suggesting that giving behavior is persistent over time. From the demographic characteristics, age and gender turn out to be significant, with individuals in the highest age quartile and females being more likely to be strongly committed. To evaluate the out-of-sample predictive performance, we randomly split the sample into two equal-sized subsamples. We then use one subsample for the estimation and the other subsample to calculate metrics for predictive accuracy. Using a threshold of 50% for the predicted probability to classify an individual as a strongly committed type, our out-of-sample prediction is correct for 91.5% of the individuals. Because

[14] As stated before, our results are robust to relaxing this restriction. See Section 4.3 for details.

[15] The no-intervention cross-section consists of 1,194 observations. We also evaluated the performance of machine learning (ML) models that can capture more complex relationships between the donor's characteristics and her type but found that ML methods do not lead to improvements over the probit regression in terms of out-of-sample predictive performance. Moreover, all our results are robust to using a simple heuristic to define donor types (see Section 4.3 for details). We also predicted three types (donation strictly above, equal to, and strictly below suggested amount) but we ended up with very small samples for the first two types.

the successful prediction of types in the majority class ($TYPE = 0$) could mask a possible failure of our model in predicting the minority class ($TYPE = 1$), focusing on true positives provides a more conservative assessment of the model's performance in the zero-inflated data. We find that calculating the number of true positives relative to all positives still results in a share of correctly predicted types of 80.4%. Using as a metric the number of true positives over the number of predicted positives, we get a share of 74.1%. We take these metrics as evidence for a favorable out-of-sample predictive performance of our first-stage model.

Not surprisingly, given a share of donors in the no-intervention group (i.e. absent any treatment) of barely 20%, the distribution of predicted types is heavily skewed toward a weak commitment to the cause: 85.0% of donors with data on baseline income are classified as weakly committed, and 15.0% as strongly committed donors.

While we believe that the distinction between strongly and weakly committed donors is a useful refinement of the common (but somewhat coarse) warm-list versus cold-list classification, we acknowledge that our approach to single out the strongly committed types relies on the specific institutional features of our setting and thus cannot be applied generally. Of course, in other institutional environments, other ways to differentiate donors by their degree of commitment to the cause might be viable.

4. Evaluating the field experiment

4.1 Estimation

We consider three outcome variables: an indicator for individuals whose donation exceeds the personalized suggested amount, an indicator for individuals who make a donation, and the donation amount. When using either the indicator for a donation or the donation amount, we exploit the full sample of 6,433 individuals who are in one of the groups used for estimation (efficiency, impact, and control) and whose type can be predicted. The regression using the indicator for donations exceeding the personalized suggested amount uses only the sample of 3,625 individuals for whom we can also determine the personalized suggested amount in the treatment year.

Using the control group as the omitted reference category, we estimate the efficiency effect and the impact effect by OLS. We consider two specifications: First, we estimate the non-interacted model

$$y_i = \beta_0 + \beta_1 EFF_i + \beta_2 IMP_i + X_i'\gamma + \epsilon_i, \tag{2}$$

where y_i denotes the outcome of interest, EFF_i is an indicator for the efficiency treatment, IMP_i is an indicator for the impact treatment, and X_i denotes the vector of strata variables (age, gender, income bracket, indicators for donation in just one or both baseline years, and parish fixed effects). Because the strata variables include indicators for income brackets, the estimation flexibly controls for income effects. We use Huber–White robust standard errors for inference. Second, we estimate a model that interacts the treatment indicators with the predicted donor type,

$$\begin{aligned} y_i = \beta_0 + \beta_1 EFF_i + \beta_2 IMP_i + \\ \beta_3 EFF_i \times \hat{TYPE}_i + \beta_4 IMP_i \times \hat{TYPE}_i + \theta \hat{TYPE}_i + X_i'\gamma + \epsilon_i, \end{aligned} \tag{3}$$

where the indicator variable $\hat{TYPE}_i$ is derived from the out-of-sample prediction based on Equation (1). $\hat{TYPE}_i$ takes on the value one (indicating strongly committed donors) for predicted probabilities greater than 50%, and zero otherwise (weakly committed individuals). In additional analyses, we show that using an alternative (heuristic) definition of types

leaves our results unchanged (see Section 4.3 for details). Since we are mostly interested in treatment responses by donor type, we consider (3) our main specification.

For inference on the interacted model, we use a bootstrap. Because predicting donor types involves an estimation, the bootstrap encompasses both the type prediction and the estimation of treatment effects. This ensures that we take the impact of the sampling variation on the predicted type into account when deriving the standard errors of the treatment effects.

It should be noted that our estimation approach identifies intent-to-treat effects. This is because we randomize the mailing of the letters in our experiment, but we have no information on whether the recipients read the letter or not. As a result, we have to assume that the take-up of the treatment variation is imperfect. Our data do not allow us to estimate the share of recipients who read the letter.

Spillovers across treatments would downward bias any differences between the treatment and control groups. Arguably, such spillovers would most likely occur within households comprising more than one adult member of the Protestant Church. We reiterate that in order to minimize the potential for spillovers, we excluded such households from the experiment (together with church employees and priests).

4.2 Results

4.2.1 Non-interacted model

Table 3 reports the estimation results. We begin by discussing the findings for the non-interacted model in Panel A described by Equation (3). Column (1) shows how the treatments affect the likelihood that donors give more than the personalized suggested amount. In response to the efficiency treatment, donors are, on average, 2.7 percentage points more likely to donate more than the suggested amount. Given a share of 6.6% of donors in the control group whose donation exceeds the suggested amount, this is a sizable effect. We also find that, relative to the 18.5% share in the control group, the efficiency treatment does not affect the likelihood for making a donation, and that despite the higher tendency to donate more than the suggested amount, the effect on the donation amount (€6.36 on average in the control group) is not significantly different from zero. In contrast, the coefficient of the impact treatment indicator is close to zero and insignificant in all three regressions. Moreover, we can reject the hypothesis that the effects of both treatments in column (1) are equal (p-value 0.053). Taken together, the evidence from the non-interacted model suggests that on average, if the charity reduces its fundraising expenses in a way that signals an improved efficiency, it triggers a positive response among donors who become more likely to donate more than the suggested amount. In contrast, increasing the impact a donation has on the cause does not, on average, affect donor behavior.

As discussed in the following paragraphs, the average treatment effects mask considerable heterogeneity in treatment responses between weakly and strongly committed donors. Since the distribution of donor types will likely differ a lot between contexts, we do not want to put too much emphasis on the average treatment effects and consider the findings from the interacted model as our main results.

4.2.2 Interacted model

Evidence on the heterogeneity of the treatment effects with respect to the degree of commitment to the cause is reported in Table 3, Panel B. To facilitate the interpretation of the effects, Table 4 additionally displays mean outcomes in the control group separately for both donor types (strongly versus weakly committed).

The coefficients of the efficiency treatment indicator (β_1) and the impact treatment indicator (β_2) capture the treatment effects for individuals who are weakly committed to the cause. A first observation is that weakly committed potential donors do not respond at all to the efficiency treatment: in all three columns, the coefficients are small and not significantly different from zero. For the impact treatment, we even find a negative effect on the

Table 3. Responses to changes in fundraising efficiency and impact.

	Donation exceeds suggested amount (1)	Makes donation (2)	Donated amount (in €) (3)
A: Non-interacted model			
Efficiency	0.027***	−0.012	0.188
	(0.010)	(0.011)	(0.546)
Impact	0.007	−0.010	0.071
	(0.010)	(0.011)	(0.539)
Efficiency = Impact	0.053	0.906	0.834
B: Interacted model			
Efficiency	0.008	−0.010	−0.498
	(0.007)	(0.010)	(0.423)
Impact	−0.001	−0.019**	−0.443
	(0.007)	(0.009)	(0.434)
Efficiency × Strongly committed	0.134***	0.013	5.86**
	(0.049)	(0.039)	(2.57)
Impact × Strongly committed	0.018	0.015	1.779
	(0.047)	(0.039)	(2.428)
Strongly committed	0.221***	0.594***	22.31***
	(0.034)	(0.030)	(1.82)
Efficiency + Efficiency × Strongly committed = 0	0.003	0.930	0.032
Impact + Impact × Strongly committed = 0	0.706	0.923	0.571
Number of observations	3625	6433	6433
Mean outcome in control group	0.066	0.185	6.36
Controls for strata variables	Yes	Yes	Yes

Notes: The table reports the results of OLS regressions. For each outcome, the table separately reports a regression of the noninteracted model (Panel A) and the interacted model (Panel B). Column (1) uses a smaller sample as compared to columns (2) and (3) because the dependent variables can only be constructed for church members for whom we can determine the suggested donation amount implied by the income-dependent donation scheme in the treatment year. All regressions include a full set of controls for strata variables (based on age, gender, the suggested donation amount in the baseline including an indicator for missing values, and parish fixed effects). Standard errors (SEs) in parentheses. Panel A: SEs are Huber–White robust. Panel B: SEs are bootstrapped. *** and ** denote significance level at 1% and 5% level, respectively. The lines with hypothesis tests (Panel A and Panel B) report *p*-values. *Source*: Authors' calculations.

probability to make a donation. Whereas 9.1% of the weakly committed individuals in the control group make a donation (see Table 4), this share is reduced by 1.9 percentage points in the impact treatment.

Our first main insight is, thus, as follows: By reducing its fundraising expenses, the charity is unable to positively affect the donation behavior of weakly committed potential donors. Whereas signaling that the reduction in fundraising expenses has improved the charity's efficiency does not affect donor behavior at all, communicating an increased impact of donations on the cause actually crowds out the willingness to give (measured by the likelihood to donate) of at least some donors. In light of the evidence in Gneezy *et al.* (2014), the finding of crowding out among weakly committed donors is a novel and surprising result. It should be noted, however, that a possible negative effect of the impact treatment is in line with economic theory. This is because, as highlighted earlier, the intervention in the impact condition can be interpreted as an infusion of seed money that is used to cover the overhead of a fundraising drive. The classic literature on the public-goods crowding-out hypothesis (Warr, 1982; Roberts, 1984; Andreoni, 1989) has argued that, if donors have altruistic preferences and a charity receives a grant from a third party, this grant fully crowds out private donations. Hence, according to this theory, seed money should crowd out

Table 4. Donation behavior by predicted donor type in control group.

	Donation exceeds suggested amount (1)	Makes donation (2)	Donated amount (in €) (3)
A: Weakly committed			
	0.022	0.091	3.03
	(0.005)	(0.007)	(0.30)
Number of observations	1,004	1,839	1,839
B: Strongly committed			
	0.285	0.729	25.71
	(0.032)	(0.025)	(1.67)
Number of observations	200	317	317

Notes: The table documents means and standard deviations of outcomes by predicted donor type in the control group. *Source*: Authors' calculations.

donations, irrespective of how the charity uses the lead donation. Interestingly, our experimental design, which nets out other potential channels through which a seed money-induced reduction in overhead could positively affect donor behavior,[16] finds at least some support for the crowding-out hypothesis for weakly committed individuals.

We would like to note that the possibility that seed money could actually crowd out private donations has been discussed in related literature. For instance, Eckel and Grossman (2003), Karlan and List (2007), and Huck and Rasul (2011) show that using seed money to finance a linear matching scheme crowds out actual donations given, and Adena and Huck (2017) present a possible remedy. In related work, Krasteva and Yildirim (2013) discuss theoretically a possible crowding of a direct grant to a charity that works through discouraging information acquisition by uninformed donors.

Next, we turn to the evidence of strongly committed donors. The coefficients of the interactions (β_3 and β_4, respectively) capture the extent to which the treatment effects of strongly committed individuals differ from the treatment effects for weakly committed potential donors. From column (1), we see that among strongly committed donors, the efficiency treatment triggers a significant shift (relative to the weakly committed) toward giving more than the suggested amount ($\beta_3 = 0.134$). The sum of β_1 and β_3 indicates that, relative to the control group, the likelihood of strongly committed potential donors to donate more than the suggested amount increases from 28.5% (see Table 4) to 42.7%, an economically sizable effect. In contrast, the likelihood to donate among strongly committed donors is not significantly affected by the efficiency treatment (column (2)). The shift toward more generous gifts among strongly committed donors that is visible in column (1), nevertheless, translates into a significant increase in the donation amount. Relative to weakly committed potential donors, the efficiency treatment triggers an increase in the average donation of €5.86. Again, the sum of β_1 and β_3 shows that the overall increase in the donation amount among strongly committed individuals is economically significant: relative to the control group mean of €25.71 (see Table 4), donations increase by 21.1%.

Regarding the strongly committed donors' response to the impact treatment, we note from Table 3 that the respective interaction effects are not significantly different from zero. To evaluate the strongly committed donors' overall response to the impact treatment, we consider the sum of $\hat{\beta}_2$ and $\hat{\beta}_4$. For all three outcomes, the sum is not significantly different from zero.

[16] Most prominently, seed money could reveal the charity's quality (Vesterlund, 2003). We would argue that in our design, a possible quality signal transmitted by the impact treatment is netted out. After all, the letter in the control group also communicates a seed money-induced reduction in fundraising expenses.

To summarize, our second main insight is that by reducing their fundraising expenses, charities can significantly increase the average donation received from strongly committed donors. Importantly, provided that the reduction in fundraising expenses is framed as an improvement in the charity's efficiency, strongly committed donors respond positively even if the impact a donation has on the cause is held constant. In contrast, a reduction in fundraising expenses that increases the impact a donation has on the cause but does not change the charity's efficiency leaves the behavior of strongly committed donors unaffected. It is worth noting that for a donor motivated by altruism, a seed money-induced reduction in overhead is a perfect substitute for her own private contribution, and would thus crowd out rather than increase donations.[17] Hence, the absence of a response to the impact treatment implies that at the margin, strongly committed donors are motivated by joy-of-giving, or warm glow, rather than altruism. Finally, it has long been recognized that the behavior of donors is strongly affected by a preference to single out and give to high-quality charities (Vesterlund, 2003; Bekkers and Wiepking, 2011). Our findings on strongly committed donors are in line with this notion: the power of the efficiency treatment suggests that strongly committed donors respond to overhead-related information due to a preference for charities that operate efficiently.

4.3 Robustness

In the following, we briefly summarize the results of some further analyses and robustness checks.

4.3.1 Income heterogeneity

Although the treatment groups are balanced in observables (see Table 2 and Table A1 in the Supplementary Appendix), the diverging responses by strongly versus weakly committed potential donors could (partly) reflect income differences. For instance, high-income individuals could be more likely to give, and at the same time be more responsive to efficiency-related information. In contrast, low-income individuals could be less likely to give, and at the same time be more likely to have purely altruistic preferences for giving.

To explore this possibility, we run entropy-reweighted estimations of our treatment effects, following the methodology of Hainmueller (2012). The idea is to construct weights for donors classified as strongly committed ensuring that their weighted pre-treatment income distribution follows the income distribution of weakly committed potential donors. We implement the reweighting using a series of six indicators for the different income brackets used by the church to derive the personalized suggested donation amount. The weights thus adjust the income distribution of strongly committed donors in a very flexible way.

Table A4 in the Supplementary Appendix documents that all our findings are robust to the reweighting. This bolsters our confidence that the treatment effect heterogeneity between the different donor types is not an artifact driven by between-type income heterogeneity.

4.3.2 Estimation sample

Table 3 uses different samples across regressions. As a further robustness check, we repeat the estimations using the same (smaller) sample for all estimations (all donors with income information both in the baseline and in the treatment year). Table A5 in the Supplementary Appendix shows that all our findings are confirmed.

In further analyses, we tested if our decision to exclude individuals living in households with more than one adult church member affects our results. Table A6 in the

[17] Again we make the assumption here that the provision of seed money (i.e., the transfer from the state church to the church districts) does not signal the districts' quality as fundraisers. As argued before, our experimental design supports this assumption.

Supplementary Appendix shows that our results are unaffected if we relax this restriction. Table A7 in the Supplementary Appendix repeats the regressions when considering households instead of individuals. Here, we construct the outcomes and the suggested donation amounts by aggregating over all adult church members in a household. Again, all our results are confirmed.

4.3.3 Alternative definition of donor types

The analysis of heterogeneous treatment effects rests on a prediction of donor types. Table A8 shows that we obtain similar results if we use a simple heuristic instead of a model-based prediction. The heuristic defines strongly committed donors as individuals who have either donated the suggested amount in all baseline years or overpaid at least once, and weakly committed donors as individuals who gave less than the personalized suggested amount at least once and did not overpay in any year.

4.3.4 Estimations without controls

The estimations in Table 3 control for strata variables. Table A9 in the Supplementary Appendix demonstrates that without controls, we obtain very similar results.

5. Conclusion

This article asks how charities can use overhead reductions to induce giving. In a field experiment, we randomly varied the information that potential donors received about fundraising costs. The experimental design allows us to separately identify two channels through which overhead reductions could affect behavior: an efficiency channel (donors give more if the charity's fundraising becomes more efficient), and an impact channel (donors give more if the charity manages to increase the impact a donation has on the cause).

Exploring the treatment responses by the donors' degree of commitment to the cause, we obtain the following main results. First, donors who are strongly committed to the cause ignore impact-related information but respond positively to a signal of improved efficiency along the intensive margin. Specifically, they give 21.1% more on average, and their likelihood to donate more than the suggested amount increases from 28.5% to 42.7%. In contrast, the likelihood of giving among strongly committed donors is not significantly affected by a signal of improved efficiency. Donors who are only weakly committed to the cause do not react positively to any of the treatments. They do not respond at all to a signal of improved efficiency, and when learning that the impact of a donation has increased, they become less likely to give. The latter finding is consistent with the public-goods crowding-out hypothesis (Andreoni, 1989).

While we acknowledge that our context is specific, we believe that our study helps to shed light on the question why specifically donors are averse to overhead in charity. Regarding strongly committed donors, arguably the most important group of potential donors from a charity perspective, our findings suggest that overhead aversion is mainly driven by a preference to give to high-quality charities (or an efficiency motivation) rather than concerns regarding the impact a donation has on the cause.

Supplementary material

Supplementary material is available on the OUP website. These are the data and replication files and the online appendix.

Funding

No external funding was received for the study. The costs of implementing the treatments were covered by the Evangelical Lutheran Church in Bavaria.

Acknowledgements

We thank the Evangelical Lutheran Church in Bavaria and especially Johannes Bermpohl for partnership on this project. We have benefited greatly from discussions with Matteo Cervelatti, Ulrich Glogowsky, Johannes Hermle, Dominika Langenmayr, Ricardo Perez-Truglia, Sebastian Siegloch, Joel Slemrod, and seminar participants at various locations. An editor and three anonymous reviewers provided valuable suggestions. The article represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff. The experiment was approved by the ethics committee of the Department of Economics and Society at the University of Nuremberg and pre-registered at the AEA RCT Registry (AEARCTR-0001332).

References

Adena, M. and Huck, S. (2017) Matching donations without crowding out? Some theoretical considerations, a field, and a lab experiment, *Journal of Public Economics*, **148**, 32–42.

Andreoni, J. (1989) Giving with impure altruism: Applications to charity and ricardian equivalence, *Journal of Political Economy*, **97**, 1447–58.

Andreoni, J. and Payne, A. (2011) Is crowding out due entirely to fundraising? Evidence from a panel of charities, *Journal of Public Economics*, **95**, 334–43.

Bekkers, R. and Wiepking, P. (2011) A literature review of empirical studies of philanthropy, *Nonprofit and Voluntary Sector Quarterly*, **40**, 924–73.

Bittschi, B., Dwenger, N., and Rincke, J. (2021) Water the flowers you want to grow? Evidence on private recognition and donor loyalty, *European Economic Review*, **131**, 103624.

Bowman, W. (2006) Should donors care about overhead costs? Do they care, *Nonprofit and Voluntary Sector Quarterly*, **35**, 288–310.

Callen, J. L. (1994) Money donations, volunteering and organizational efficiency, *Journal of Productivity Analysis*, **5**, 215–28.

Caviola, L., Faulmüller, N., Everett, J. A., Savulescu, J., and Kahane, G. (2014) The evaluability bias in charitable giving: Saving administration costs or saving lives? *Judgment and Decision Making*, **9**, 303–16.

Dwenger, N., Kleven, H., Rasul, I., and Rincke, J. (2016) Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany, *American Economic Journal: Economic Policy*, **8**, 203–32.

Eckel, C. and Grossman, P. J. (2003) Rebate versus matching: Does how we subsidize charitable contributions matter? *Journal of Public Economics*, **87**, 681–701.

Gneezy, U., Keenan, E. A., and Gneezy, A. (2014) Avoiding overhead aversion in charity, *Science (New York, NY)*, **346**, 632–5.

Hainmueller, J. (2012) Entropy balancing: a multivariate reweighting method to produce balanced samples in observational studies, *Political Analysis*, **20**, 25–46.

Huck, S. and Rasul, I. (2011) Matched fundraising: Evidence from a natural field experiment, *Journal of Public Economics*, **95**, 351–62.

Huck, S., Rasul, I., and Shephard, A. (2015) Comparing charitable fundraising schemes: Evidence from a natural field experiment and a structural model, *American Economic Journal: Economic Policy*, **7**, 326–69.

Karlan, D. S. and List, J. A. (2007) Does price matter in charitable giving? Evidence from a Large-Scale natural field experiment, *American Economic Review*, **97**, 1774–93.

Krasteva, S. and Yildirim, H. (2013) (Un)informed chartitable giving, *Journal of Public Economics*, **106**, 14–26.

Landry, C. E., Lange, A., List, J. A., Price, M. K., and Rupp, N. G. (2010) Is a donor in hand better than two in the bush? Evidence from a natural field experiment, *American Economic Review*, **100**, 958–83.

List, J. A. and Lucking-Reiley, D. (2002) The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign, *Journal of Political Economy*, **110**, 215–33.
Meer, J. (2014) Effects of the price of charitable giving: Evidence from an online crowdfunding platform, *Journal of Economic Behavior & Organization*, **103**, 113–24.
Okten, C. and Weisbrod, B. A. (2000) Determinants of donations in private nonprofit markets, *Journal of Public Economics*, **75**, 255–72.
Roberts, R.D. (1984) A positive model of private charity and public transfers, *Journal of Political Economy*, **92**, 136–48.
Steinberg, R. (1986) Charitable giving as a mixed public/private good: Implications for tax policy, *Public Finance Quarterly*, **14**, 415–31.
Tinkelman, D. and Mankaney, K. (2007) When is administrative efficiency associated with charitable donations, *Nonprofit and Voluntary Sector Quarterly*, **36**, 41–64.
Trussel, J.M. and Parsons, L.M. (2007) Financial factors affecting donations to charitable organizations, *Advances in Accounting*, **23**, 263–85.
Vesterlund, L. D. (2003) The informational value of sequential fund-raising, *Journal of Public Economics*, **87**, 627–57.
Warr, P. G. (1982) Pareto optimal redistribution and private charity, *Journal of Public Economics*, **19**, 131–8.