WZB

Wissenschaftszentrum Berlin
für Sozialforschung

Maja Adena
Anselm Hager

Does online fundraising increase charitable giving? A nationwide field experiment on Facebook

Discussion Paper

SP II 2020–302r2

February 2020 (2nd revision February 2024)

Research Area
Markets and Choice

Research Unit
Economics of Change

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Affiliation of the authors:

Maja Adena, WZB (maja.adena@wzb.eu)

Anselm Hager, Humboldt Universität zu Berlin

Abstract

Does online fundraising increase charitable giving? A nationwide field experiment on Facebook*

Does online fundraising increase charitable giving? Using the Facebook advertising tool, we implemented a natural field experiment across Germany, randomly assigning almost 8,000 postal codes to Save the Children fundraising videos or to a pure control. We studied changes in the donation revenue and frequency for Save the Children and other charities by postal code. Our georandomized design circumvented many difficulties inherent in studies based on click-through data, especially substitution and measurement issues. We found that (i) video fundraising increased donation revenue and frequency to Save the Children during the campaign and in the subsequent five weeks; (ii) the campaign was profitable for the fundraiser; and (iii) the effects were similar independent of video content and impression assignment strategy. However, we also found some crowding out of donations to other similar charities or projects. Finally, we demonstrated that click data may be an inappropriate proxy for donations and recommend that managers use careful experimental designs that can plausibly evaluate the effects of advertising on relevant outcomes.

Keywords: Charitable giving, field experiments, fundraising, social media, competition.

JEL classification: C93, D64, D12

* We thank Steffen Huck, Marrit Teirlinck, Raphael Epperson, and the participants in the BBE Workshop (2019), the BSE Micro Workshop (2019), Groningen Cause Marketing Workshop (2019), DICE Research Seminar (2019), the CESifo Workshop on Economics of Digitalization (2019), European Economic Association Congress (2020), Society for the Advancement of Behavioral Economics Annual Conference (2020), Recent Advances in the Economics of Philanthropy Workshop (2021), ZEW Mannheim Research Seminar (2022), Magdeburg Research Colloquium (2022), Behavioral Economics Design Initiative Conference (2022), Seminar at the University of Hamburg (2023), 15th Nordic Conference on Behavioral and Experimental Economics (2023), 2023 Stockholm School of Economics Brown bag seminar, GfeW annual meeting (2023) and many others for their helpful suggestions and comments. We are grateful to Julian Harke, Katharina Dorn, Meret Borchmann, Max Padubrin, Steffen Mayer, Marius Werz, and Lena Simmat for their excellent research assistance. We thank Save the Children for their cooperation and Aktion Deutschland Hilft and betterplace.org for providing the data. Maja Adena gratefully acknowledges financial support from the German Research Foundation (DFG) through project number 417014946 and CRC TRR 190 (project number 280092119). The design for this experiment was preregistered at EGAP (https://osf.io/9j4g6) before the data from the experiment was made available to the researchers.

1 Introduction

Online advertising is becoming an ever more important tool for fundraisers. In the United States, the share of online giving has been rising in the last years, reaching 8.5% of all donations in 2018[1] and can be expected to double by 2025.[2] While there are a few studies on online advertising effectiveness in the for-profit market,[3] the question of online fundraising effectiveness has received little systematic examination. Information on the nonprofit market predominantly consists of anecdotal evidence, fundraisers' intuition, and advice from for-profit consultancies (Landry et al., 2006, 2010). Yet bad decisions about fundraising expenditures not only affect charities' finances today but also impact future willingness to give to such charities (Gneezy et al., 2014) and even trust in the nonprofit market as a whole (Adena, 2016).

Existing studies on online fundraising, starting with Chen et al. (2005), have typically been limited to one clearly defined environment, such as a single donation platform.[4] Such designs suffer from several difficulties. First, they are plagued by very low statistical power because donations are infrequent and volatile.[5] Second, when donors give via a link embedded in the ad, they may simply be substituting away from other

[1] https://institute.blackbaud.com/the-blackbaud-institute-index/ (viewed on August 12, 2019). This figure is similar for the UK (8.4%, https://www.nptuk.org/philanthropic-resources/uk-charitable-giving-statistics/, viewed on August 12, 2019) and Germany (9%, https://www.betterplace.org/c/neues/online-fundraising-auf-betterplace-org-das-jahr-2016-in-zahlen, viewed on August 12, 2019).

[2] Assuming the constant growth rate of additional 1.2 percentage point annually as suggested by the Blackbaud Institute, see footnote 1.

[3] See, for example, Lewis et al. (2015) and the references cited therein. For studies on online advertisement effectiveness in the context of voting, see Bond et al. (2012) and Hager (2019).

[4] In Chen et al. (2005) the researchers observed button clicks and direct donations. Castillo et al. (2014) asked donation platform users to post solicitation messages on their Facebook walls or as direct messages, observed whether a message had been posted, and traced whether a hyperlink in a post had been clicked and a donation made. On the platform JustGiving, Bøg et al. (2012) studied how donors reacted to donations already made. On the DonorsChoose platform, Meer (2017) studied how matching grants for certain projects affected donations to other projects. Scharf et al. (2022) studied responses to major donation appeals for donors who had an account administered by the Charities Aid Foundation. All of those studies observed behavior of a narrowly specified group and only within the studied environment.

[5] For example, Chen et al. (2005) observed 24 donations after more than 150,000 impressions. Castillo et al. (2014) traced five donations in response to friends' Facebook wall posts or private solicitation messages. See Lewis and Rao (2015) for a discussion of the power problem in the context of commercial advertising.

donation channels (Blake et al., 2015) or from giving at some other time (Adena and Huck, 2019). Third, the opposite is also possible: Online ads may lead ad recipients to take an action at a later stage or via a different channel, which the researchers do not observe (Lewis and Reiley, 2014). Finally, such designs cannot observe general equilibrium effects, including potential crowding out of donations from competitors.

The present study overcomes these challenges by administering an unusually large geo-randomized online experiment in conjunction with a charity, namely Save the Children. We randomly assigned 94% of Germany's 8,181 postal codes (*Postleitzahl*, or PLZ) to a 14-day Facebook campaign of Save the Children fundraising videos or to a pure control group. Our main outcome is Save the Children's full universe of donations at the postal-code level. By studying changes in the overall donation revenue by postal code across all possible donation channels, the design thus circumvents the aforementioned difficulties and bypasses channel-substitution and measurement issues. By studying almost all of Germany's postal codes across a period of 12 weeks and using a largely untargeted campaign, the design enhances statistical power and ensures a high degree of external validity with respect to online fundraising. Moreover, our design allows us to discuss general equilibrium effects because the experiment covered an entire country and a large portion of the population. We are therefore in a position to address questions concerning potential effects on competing charities and spillovers (Banerjee et al., 2017a,b).

The results show that the largely untargeted fundraising campaign increased total donation revenue and donation frequency to Save the Children during and in the five weeks after the campaign. The increase in donation revenue is estimated to be €17.65 per million inhabitants per day from the average of €129.3 in the control group, while the increase in frequency is estimated to be 0.21 donations per million inhabitants per day from the control-group average of 1.80. Those point estimates translate into €1.45 in additional donations for each €1 spent in immediate returns. Assuming a realistic long-term multiplier for a new donation of 1.75,[6] this implies a return of €2.53 in the

[6]See Section 4.3 for details.

long term per €1 initially spent making the campaign profitable for the fundraiser. Importantly, the increase is not the result of a substitution between different donation channels to the same charity because our data accounts for all donations made to Save the Children. It is also not the result of intertemporal substitution, given that we accounted for donations during a sufficiently long period after the campaign. The latter results emphasize the long-lived nature of the effects of advertising (Lewis and Reiley, 2014).

Importantly, using data on charitable giving to other similar charities and projects, we find evidence that the Save the Children campaign led to some substitution away from similar causes. This suggests that donors may not approach their budgets for charitable giving with the degree of flexibility suggested in some previous research (Meer, 2017; Donkers et al., 2017; Gee and Meer, 2019; Grieder and Schmitz, 2020; Deryugina and Marx, 2021; Gallier et al., 2023; Jayaraman et al., 2023). Rather, fundraising campaigns seem to prompt individuals to shift their donation expenditures between charities. This implies that charities are competing for scarce resources (Rose-Ackerman, 1982; Reinstein, 2011; Reinstein and Riener, 2012; Bilodeaua and Slivinski, 1997; Lacetera et al., 2012; Petrova et al., 2024) rather than acting as complements (Krieg and Samek, 2017; Lange and Stocking, 2012; Filiz-Ozbay and Uler, 2019).

In order to uncover the mechanisms behind the increase in giving to Save the Children, we implemented a 2x2 factorial design in the treatment group. First, we randomized whether the video was designed to induce empathy for those in need or whether it was intended to highlight the effectiveness of the organization. Second, we randomized whether Facebook's algorithm was free to decide how advertising dollars were allocated across treated postal codes or whether we assigned a fixed budget to each postal code proportional to the estimated donor potential and Facebook reach. The empathy video attracted more attention and more frequent immediate donations but the long-term differences were not significant. Compared to the fixed postal-code-level budgets, the treatment that allowed Facebook to distribute impressions freely between postal codes led to higher donation frequency and values, especially in the short term.

While any conclusions are necessarily limited to the specific implementation of our campaign, we interpret these results on the additional treatment variation as lending external validity to our main results—that no matter the specific campaign design, online fundraising works—and as supporting the existence of the "power of asking" (Yörük, 2009; Andreoni and Rao, 2011) in an online context despite clearly reduced social pressure.

Finally, we document that relying on intermediate metrics like click-through ratios and time spent watching videos might lead to conclusions that do not necessarily align with results based on long-term measures. We therefore advise advertising and fundraising managers to use careful experimental designs that rely on relevant outcomes and can account for substitution and long-term effects.

The rest of the paper proceeds as follows. In Section 2, we introduce the experimental design. In Section 3, we analyze the effects of Facebook video fundraising on giving behavior and study the effects of Save the Children video fundraising on competing charities. In Section 4, we examine potential differences on two dimensions: (i) between two types of videos and (ii) between two degrees of control over the Facebook algorithm regarding the distribution of impressions between postal codes. We also provide robustness checks, additional analyses, and a discussion of our main effects including an evaluation of profitability from the perspective of the fundraiser. Section 5 concludes.

2 Design

We partnered with one of the world's largest charities, Save the Children, in order to test the effectiveness of online fundraising. The fundraising campaign took the form of a video advertisement on Facebook. The gross sample in this experiment consisted of all 8,181 German postal codes, all of which can be targeted via Facebook's advertising manager.[7] For each postal code we knew Facebook's estimated reach, that is, the

[7]Facebook's targeting procedure relies on a variety of data sources, including GPS signals, IP addresses, and individual-level data. If this information is noisy then our results can be interpreted

number of individuals Facebook estimates it can target. We excluded the lower 5% as well as the upper 1% of the reach variable for several reasons. First, since half of the fundraising budget was distributed proportional to the reach variable, we needed to avoid falling below Facebook's minimum advertising spend in small postal codes and overspending in very large ones that would jeopardize our budget.[8] Second, we wanted to avoid having an overly high advertising spend in those postal codes with the highest reach, as this could have given rise to significant spillover concerns.[9] Moreover, we considered these types of outliers to be a threat to covariate balance. On the other hand, we needed to keep the final sample as large as possible for power reasons. The final number of postal codes was 7,686.

By choosing geographical areas instead of individuals as the unit of analysis, we sought to overcome the following challenges inherent in individual-level online experiments: (i) Tracing individuals is never an exact science, and those who can be traced for longer periods of time likely differ from the general population. (ii) Matching traced individuals to donations through other channels and later donations, especially offline, is oftentimes not possible, although this information is crucial in order to estimate the total effect of any advertising or fundraising campaign. (iii) Charitable giving in response to untargeted online fundraising is a low frequency behavior. (iv) Keeping the control and treatment groups comparable in individual-level experiments requires posting unrelated ads for the control group, which is costly.[10]

To ensure balance across pretreatment variables, we relied on a machine learning technique of gradient boosting to build a targeting model for all postal codes. The model predicted future donations based on past donations[11] and other salient pretreat-

as lower bound estimates. We will address this issue later on.

[8]Half of the treated postal codes was assigned to a treatment with fixed postal-code budgets. The budgets were assigned proportionally to Facebook reach and estimated potential. Facebook requires a minimum spend of €1 per day.

[9]See the discussion on spillovers in Section 4.3.3.

[10]Without unrelated ads, more active individuals are more likely to receive an ad, that is, to end up in the treatment group, but also more likely to be active in all online contexts, including online giving (activity bias, see Johnson et al., 2017).

[11]All donation data provided to us were anonymized and aggregated at the PLZ-day level such that no conclusions can be drawn about individual persons.

ment postal-code characteristics, including socio-demographic and political variables.

We multiplied this donation potential with Facebook's estimated reach, sorted the postal codes in descending order according to this variable, and assigned each of the six consecutive postal codes to one block. In any given block, we randomly assigned the postal codes to one of the following conditions: two postal codes received no ads (the control group) and four postal codes were allocated to the ad condition (the treatment group). In the treatment group, in each block, postal codes were further randomly assigned to one of four treatments following a 2x2 design: one of two video types and one of two impression allocation strategies. One video was designed to induce empathy with those in need (empathy video), while the other was designed to highlight the effectiveness of the organization (effectiveness video). In addition, we randomized whether Facebook's algorithm was free to decide how advertising spend was allocated across postal codes (free allocation) or whether we assigned a fixed budget to each postal code—proportional to estimated donor potential and Facebook reach (fixed postal-code budgets). We did not implement any further targeting beyond the postal-code level. More specifically, there was no targeting at the individual level. In Table A1 in the appendix, we show that, for the available baseline characteristics of the postal codes, there were no differences between the treatment groups.[12] Figures A1, A2, and A3 in the appendix show the spatial distribution of treatments. The design for this experiment was preregistered at EGAP registry (number blinded) before the data from the experiment was made available to the researchers.[13]

The natural field experiment was implemented between November 10[14] and 23, 2017. This is a typical time of the year for charities in Germany to run fundraising campaigns. The treatment length of 14 days was similar to the median duration of all for-profit campaigns studied by Lewis and Rao (2015). For our analysis, we used

[12] Out of 39 presented t-tests only one is significant at $p<0.05$ and one at $p<0.1$, both for the difference between the empathy and effectiveness video groups, which is well within the margin of statistical error.

[13] There were some changes to the preregistered design. For blocking we additionally used the reach variable, and the second treatment dimension regarding impression allocation strategy was added.

[14] In the evening hours.

daily postal-code-level donation data from October 10 to December 31, 2017, thus 31 days before the campaign (pretreatment period) and 38 days after the campaign (posttreatment period).[15] The posttreatment period was a little longer than the 1–4 weeks used in Lewis and Rao (2015), which they described as standard in the for-profit industry. However, in the nonprofit sector, the bulk of donations arrive around Christmas time,[16] before the end of the fiscal year, which is December 31 in Germany. Therefore, we expected the treatment effect to be relevant when those donation decisions were being made but to die out in the new year.[17] Note that the specific timing of the experiment provides an important test for the intertemporal substitution: If people had planned to donate to Save the Children in December and received an ask on Facebook in November, they might have decided to respond immediately instead of waiting until later. On the other hand, this period of time is a good test for long-term effects as well. People who made their decision to donate in December might still have remembered the Save the Children ad and have directed their donations to that charity.[18]

The fundraising ad appeared in users' Facebook news-feeds in between posts from friends, and other advertisers. It included a subtitled video embedded into a larger banner with the Save the Children logo. If not disallowed in the individual's account and device settings, once the user scrolled to the video, it began playing (with or without sound) until the user scrolled away. The user could click on the video to see it in a larger format and could also click on a button forwarding them to the Save the

[15] As preregistered. In fact, we have data for the first 10 days of January 2018 and use them in robustness checks in section 4.3.

[16] In 2017 in Germany, the total donation revenue to all charitable organizations in December amounted to 20% of that for the whole year while this number was 32% for November and December together (https://www.spendenrat.de/wp-content/uploads/Downloads/Bilanz-des-Helfens/bilanz-des-helfens-2018-deutscher-spendenrat.pdf, viewed on November 18, 2021). In the US, donations in December account for 17.5% of those made across the whole year, while donations made during the "giving season" between Thanksgiving and Christmas account for 33.6% (Müller and Rau, 2019).

[17] The decision to exclude the period in the new year follows the rule from Lewis et al. (2015) to exclude weeks in which the expected effect is less than one half of the average effect over all previous weeks. Note that in our data, the level of giving in the first 10 days in January is 3:10 compared to the last 10 days in December.

[18] Unfortunately, if both effects are at play at the same time, they might cancel each other out, but at least we can evaluate the total long-term effect.

Children website.

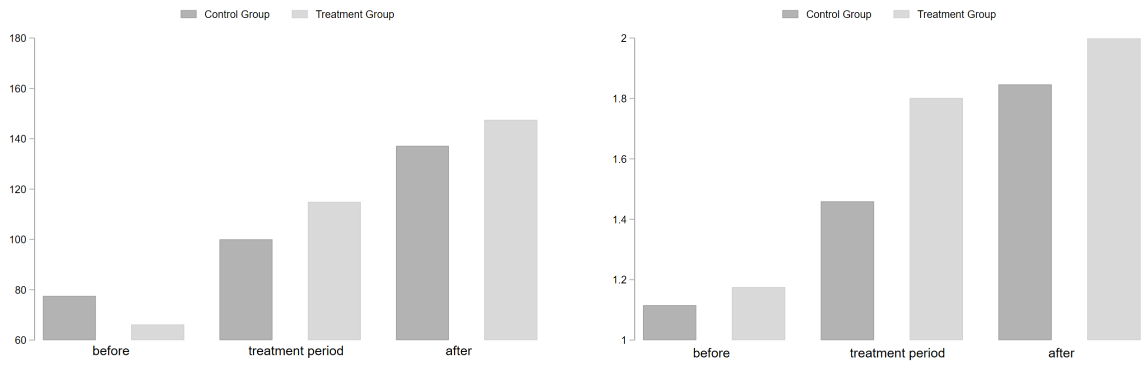
3 The effects of Save the Children Facebook video fundraising

The total number of impressions was more than 2.25 million presented to 1.9 million people in the treated postal codes. The total number of people that Facebook purported to be able to reach in the treated postal codes was 19 million and the total population (including children) in the treated postal codes is 52 million. This means that in treated postal codes, every tenth Facebook user received an impression of the video at least once. On more than 500,000 occasions the video ran for at least three seconds. In more than 16,200 instances users clicked on the video and in over 1,500 instances they clicked on the forwarding button. In the period under study, Save the Children received 13,269 individual donations that could be linked to postal codes totaling almost €1 million in giving. The data provided to us were aggregated at the PLZ-day level. There were 11,140 nonzero PLZ-day donations, and half of the postal codes received at least one positive donation. The most frequent donations were of €10 followed by €5, €50, and €100. The average donation was €87, and the median was €30. There were 68 PLZ-day observations greater than €1,000. From this point on, we winsorize the PLZ-day level donations at €1,000 in order to reduce the influence of outliers and to reduce variance.[19] For each postal code, we aggregated donation revenue and the number of donations at the period level: before, during, and after the treatment (or during together with after the treatment). Finally, we normalized those variables by population size and period length so that our outcome variables measure donation revenue and frequency per million inhabitants per day.

[19]This is standard in the literature (see, for example, Kessler and Milkman, 2018). Unfortunately, we cannot winsorize at the individual level. Yet, in 56 cases out of the 68 affected PLZ-day donations only one person donated and in 12 cases two people donated. There is no meaningful difference in the estimates if we do not winsorize, although greater variance in the outcome variable affects statistical precision.

Figure 1. Average Outcomes Before, During, and After the Treatment

Donation revenue	Donation frequency
Per million inhabitants per day	



Notes. Averages over 7,686 postal codes. Pretreatment period (before): 31 days. Treatment period: 14 days. Posttreatment period (after): 38 days.

3.1 Main effects

Our unit of observation is defined by postal codes, with donation revenue and donation frequency per million inhabitants per day serving as the primary outcome variables. To offer a preliminary insight into the data, Figure 1 presents an overview. The left panel illustrates the average donation amount per million inhabitants per day across three distinct periods—before, during, and after the treatment—and grouped by treatment status. While in the pretreatment period the average donation amount was slightly smaller in the treatment group than in the control group, it increased during the treatment and posttreatment periods. The right panel shows the average number of donations per million inhabitants per day during each of the three periods in a similar manner. While the average number of donations was slightly higher in the treatment group before the experiment, this difference was much larger during the campaign and somewhat larger after the campaign. In both panels, we observed an increase in giving over time consistent with Christmas and end-of-fiscal-year effects. Table A2, Panel A in the appendix provides summary statistics by period.

Before proceeding to the main analysis, we tested for the existence of pretreatment differences between the treated and untreated postal codes. Table 1 presents the results of linear regressions, both with and without control variables, using donation revenue (frequency) in the pretreatment period as the outcome variable. We can confirm that there were no statistically significant pretreatment differences between the treatment and the control group in terms of donation level or frequency.[20]

Table 1. Pretreatment Differences in Donations to Save the Children

Dependent variable:	Per million inhabitants per day			
	Donation revenue		Donation frequency	
	(1)	(2)	(3)	(4)
Video fundraising	-11.262	-11.051	0.060	0.068
	(13.672)	(13.473)	(0.069)	(0.068)
Controls		yes		yes
Randomization blocks FEs		yes		yes
Observations	7,686	7,686	7,686	7,686
R^2	0.000	0.167	0.000	0.186

Notes. Linear estimations in Stata. Controls include: population, shares employed and Catholics, and the number of post codes per county. FEs: fixed effects. Pretreatment period: 31 days. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Considering the randomized experiment, the assessment of the causal effect of the video fundraising campaign on donation revenue and frequency is straightforward—simply by comparing donations in treated and untreated postal codes. This comparison was conducted using linear regressions, the results of which are presented in Table 2. We studied both the immediate effect during the two weeks of the campaign (short term, Panel A) and the more comprehensive effect, which is the combined effect of during and after the campaign until the end of the year (long term, Panel B). Columns (1) and (2) present results for donation revenue, while Columns (3) and (4) present results for donation frequency. In Columns (2) and (4), we included control variables such as the lagged dependent variable, randomization blocks fixed effects, and a few other controls that helped to (minimally) increase precision. This is our preferred specification. While the short-term effect of video fundraising on donation revenue is not significant, the long-term effect is significant at p<0.1. The coefficients

[20]This also holds for more extensive sets of control variables; not presented here.

Table 2. Effects of Video Fundraising on Donation Level and Frequency

Dependent variable:	Per million inhabitants per day			
	Donation revenue		Donation frequency	
	(1)	(2)	(3)	(4)
Panel A: Short term				
Video fundraising	14.966	15.734	0.344***	0.331***
	(11.982)	(12.123)	(0.125)	(0.125)
Controls		yes		yes
Randomization blocks FEs		yes		yes
Observations	7,686	7,686	7,686	7,686
R^2	0.000	0.176	0.001	0.189
Panel B: Long term				
Video fundraising	15.970*	17.652*	0.219***	0.211***
	(9.678)	(9.630)	(0.077)	(0.073)
Controls		yes		yes
Randomization blocks FEs		yes		yes
Observations	7,686	7,686	7,686	7,686
R^2	0.000	0.205	0.001	0.283

Notes. Linear estimations in Stata. Controls include: lagged dependent variable, population, shares employed and Catholics and the number of post codes per location. Short term: effect during the campaign (14 days). Long term: Effect during and after the campaign (52 days). Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

are similar in magnitude and suggest an additional €15–18 in donations per million inhabitants per day. The coefficients in the regressions with donation frequency as the outcome variable are highly significant and suggest additional 0.34 donations per million inhabitants per day in the short term and additional 0.22 donations in the long term.

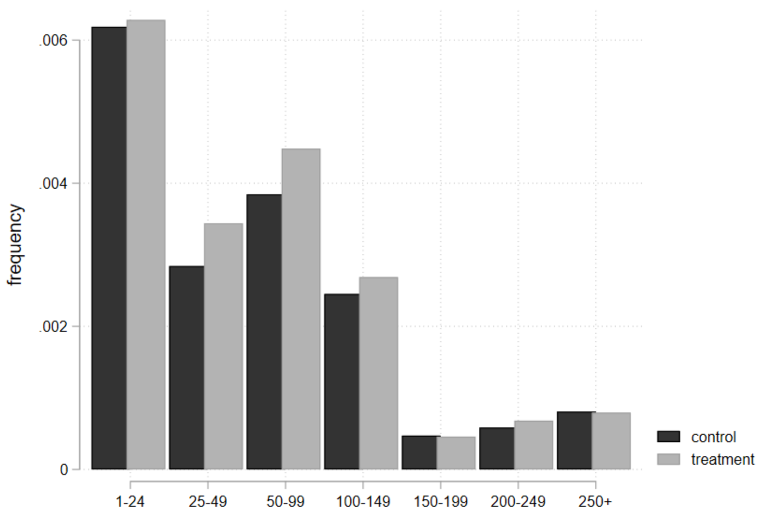
The inclusion of a lagged dependent variable (and other controls) leads to a slight increase in the coefficient in the revenue specification and a slight decrease in the frequency specification, which directionally corrects for the pretreatment imbalances, which, although nonsignificant, are present. Adding further control variables has a minimal impact on the results (not presented here). However, if we are concerned that the (nonsignificant) pretreatment differences are stable differences, we might prefer a difference-in-difference (DiD) specification. We present results of such specifications in the appendix, Table A3, and confirm that the significance levels and conclusions remain unchanged.[21]

The positive effects on donation frequency suggest that online fundraising primarily generated new donations rather than increasing the amount contributed by those who would have given regardless. Figure 2, which confirms this intuition, shows frequencies of PLZ-day donations in the treatment period by treatment status, with the zero category being the omitted category. It suggests that there were additional donations in the range of €25–€149 rather than a shift in the number of donations from lower to higher categories in the treatment group.

The results in Panel B of Table 2 suggest a long-term positive effect of the fundraising campaign and provide evidence countering (any sizable) intertemporal substitution. Of course, both effects could be at play and might have partly canceled each other out. Overall, we conclude that online fundraising has a causal effect on additional donations and that these additional donations cannot be attributed to any substitution of donations within the same organization, either regarding donation channel—because

[21] More specifically, in the appendix Table A3, we regress $Y_t - Y_{t-1}$ on $T_t - T_{t-1}$, with Y being the outcome variable, t the time index, T the treatment dummy, and T_{t-1} being always equal to zero.

Figure 2. Frequencies of Different Donation Values in the Treatment Period



Notes. The available data is aggregated at the PLZ-day level. In the treatment period, there are two donations in 3.5% of instances, three in three instances, and four donations in one instance. In those cases, we assign the average donation to the respective category. Zero-donations are the omitted category.

we account for total donations—or time frame—because we account for a sufficiently long period after the campaign.

3.2 The effects on the competition

The design of our field experiment allowed us to study the effects of Save the Children video fundraising on donations to other charities. We obtained data on other charities from two different sources.

The first source is an alliance uniting 23 charities that are active in similar domains, including humanitarian help, international relief, and support for children. The data only include online giving, but the total donation revenue over the period studied was four times that of Save the Children, and the share of postal codes with positive donations was greater than 70% (see Table A2, Panel B in the appendix for descriptive statistics of the data). As before, we first winsorized the PLZ-day level donations at €1,000. Then, we aggregated donation revenue and the number of donations for each postal code at three periods: before, during, and the combined long term, which includes both the during and after the treatment periods. Finally, we normalized those variables by population size and period length so that our outcome variables measure donation revenue and frequency per million inhabitants per day.

Before proceeding to the analysis of the treatment effect on competition, we test the pretreatment balance in terms of the outcome variables. In fact Table A4 in the appendix, which is equivalent to Table 1, shows that there are significant pretreatment differences. Therefore, simple cross-sectional estimates would provide biased estimates. We address this issue in two ways. First, as in Table 2, Columns (2) and (4), we control for the pretreatment levels of the outcome variable. The results can be found in Table 3, Columns (1) and (3). Second, in Columns (2) and (4), we regress the treatment dummy on the difference between average daily donations (frequency) in the treatment period and the period before. Note that this is equivalent to a DiD estimation. Panel A presents the short-term results, that is, the effect during the fundraising campaign, while Panel B presents the long-term results that combine the campaign period and the posttreatment period. In Columns (1) and (3), the dependent variable is donation revenue per million inhabitants per day, and in Columns (2) and (4), it is donation frequency per million inhabitants per day. The results of the regressions suggest that Save the Children fundraising reduced revenue to the other 23 charities by almost €62–90 per million inhabitants per day in the short term and by €25–56 in the long term. The DiD estimate is significant at $p<0.1$. The effect on donation frequency was a (nonsignificant) reduction in the number of donations by 1.240–1.396 in the short term and by 0.136–0.306 in the long term.

The second source of data on other charities is the largest German donation platform: betterplace.org (see, for example, Altmann et al., 2018; Jayaraman et al., 2023, for a description of the data). On this platform, potential donors can contribute to different projects (charities can present several projects), which are tagged with different (usually multiple) categories like children, animals, refugees, development, sports, religion, and so on. The data that we received exclude donations to projects by Save the Children and is aggregated at the PLZ-day-project level. For each PLZ-day-project

Table 3. Effect of the Save the Children Campaign on Donations to 23 Similar Charities

Dependent variable:	Per million inhabitants per day			
	Donation revenue	Δ donation revenue	Donation frequency	Δ donation frequency
	(1)	(2)	(3)	(4)
	Panel A: Short term			
Video fundraising	-61.622	-91.101*	-1.240	-1.169
	(53.192)	(51.861)	(1.214)	(1.059)
Controls	yes		yes	
Randomization blocks FEs	yes		yes	
Observations	7,686	7,686	7,686	7,686
R^2	0.198	0.000	0.170	0.000
	Panel B: Long term			
Video fundraising	-24.893	-59.193*	-0.136	-0.255
	(29.543)	(32.962)	(0.347)	(0.318)
Controls	yes		yes	
Randomization blocks FEs	yes		yes	
Observations	7,686	7,686	7,686	7,686
R^2	0.254	0.000	0.206	0.000

Notes. See notes to Table 2. Δ is the difference between the average per million per day donation revenue (frequency) in the treatment period (treatment and post-treatment in Panel B) and in the pretreatment period: $\Delta Y = Y_t - Y_{t-1}$.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

entry, we know the donation sum, the number of donations, the anonymized charity and project ID, and the project category tags. We used those category tags to divide the projects in two separate groups: children-related projects (35,434 single donations totaling €2,525,019 to 1,543 distinct organizations and 1,847 projects) and other projects (60,864 single donations totaling €3,399,854 to 2,010 distinct organizations and 2,595 projects). In total, the volume of giving was much higher than for Save the Children for the same period of time, and we observed non-negative giving in more than 80% of postal codes (see Table A2, Panel C in the appendix for descriptive statistics of the data). As before, we first winsorized the PLZ-day-project level donations at €1,000. Then, we aggregated donation revenue and the number of donations for each postal code at three periods (and by project type): before, during, and the combined long term, which includes both the during and after the treatment periods. Finally, we normalized those variables by population size and period length so that our outcome variables measure donation revenue and frequency per million inhabitants per day. Table A4 in the appendix shows that there were no significant pretreatment differences. Given two types of projects (children related and not children related) per postal code, we included interaction of the treatment dummy with the children-related type of project and also included postal-code fixed effects. Table 4 presents the results. We find a negative interaction effect in all cases but it is only significant for the donation revenue in the short term at $p<0.05$.[22] This weakly suggests that video fundraising by Save the Children drained donation money from other projects that benefited children at Betterplace.

Of course, we do not have data on all competitors but, based on two separate pieces of evidence, it is reasonable to assume that any further effects should go in the same direction.[23] Our results suggest that the Save the Children fundraising campaign may

[22] A regression without postal-code fixed effects but with control variables as in Table 2 leads to the same conclusions. In a DiD specification all interaction coefficients are highly significant, see Table A5.

[23] Note that in both sets of data there is an overlap and that the effects cannot simply be added: The charity alliance collects donations on betterplace.org but also via other online channels while other charities are also active on betterplace.org. The available data do not allow us to remove this overlap. Still, even taking one or the other data source, the magnitudes of the effects on the competition seem

Table 4. Effect of the Save the Children Campaign on Donations to Projects on betterplace.org

	Short term		Long term	
Dependent variable:	Per million inhabitants per day			
	Donation revenue	Donation frequency	Donation revenue	Donation frequency
	(1)	(2)	(3)	(4)
Video fundraising x children related projects	-40.637**	-0.241	-9.447	-0.198
	(20.248)	(0.310)	(25.356)	(0.249)
Children related projects	-32.010*	-1.109***	-96.458***	-2.458***
	(16.933)	(0.244)	(19.968)	(0.207)
Controls	yes	yes	yes	yes
Randomization blocks FEs	yes	yes	yes	yes
Project types	2	2	2	2
Observations	15,372	15,372	15,372	15,372
R^2	0.598	0.730	0.741	0.747

Notes. See notes to Table 2. Controls include: donations or frequency in the period before, post code fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

have reduced donations to similar causes, implying a substitution effect. While there are papers that have studied substitution over time (Adena and Huck, 2019) or over many charities (Meer, 2014), we were able to provide an experimental setting that allows us to study both together (see also Scharf et al., 2022).

4 Further results

4.1 Video content

A traditional view of advertising is that it provides relevant knowledge that informs decisions of individuals. In the context of fundraising, this knowledge could include information about the neediness of certain individuals or groups, how donations will be used by a charity, and what donations can achieve. In practice, the informational content of many advertisements and donation asks is limited. For example, most consumer ads do not provide price information, and most donation asks do not state how much relief a donation will buy. Rigorous field experiments on ad content for consumer goods include Bertrand et al. (2010), who varied several content characteristics.

to be larger than those on Save the Children, though the estimates are subject to large confidence intervals. If the effect of the campaign on competing charities is indeed higher than the effect on Save the Children, this could be explained in at least two ways: (i) The Save the Children campaign may have displaced online fundraising efforts on Facebook by other charities, increasing the likelihood of their ads appearing in the control group. (ii) Recurring donors to other charities may have switched to Save the Children and opted to make lower donations to the latter.

Examples in research on charitable giving include laboratory experiments by Eckel et al. (2007) on information overload and Andreoni (1995) on positive versus negative framing.

For our test of the effects of video content, we chose two types of videos: one designed to activate feelings of empathy and another stressing the competence and effectiveness of the organization.[24] We chose both types of video based on relevant research in the field, a discussion of which can be found in the appendix B.

In Table 5, Panel A, we present the results of the regressions similar to those in Table 2 (with controls) but now differentiate between the empathy and effectiveness videos. In the short term, Column (1) and (2), we see that the coefficients on the empathy video treatment are much larger than those on the effectiveness video. They are significant at $p<0.1$ (revenue) and $p<0.01$ (frequency), while the coefficients on the effectiveness video are much smaller and nonsignificant. However, the coefficients are not statistically different from each other.[25] While, in the short term, the empathy video seems to be more effective, this is no longer the case in the long term; the differences between coefficients are smaller and the effectiveness video coefficients are now significantly different from zero. We conclude that there are no differences in the effects by video type (a difference may exist but we are underpowered to detect it).

In Table 6, Columns (1) and (2), we present some intermediate metrics that point to the mechanism behind the effects of both video types (see Table A6, Columns (1) and (2) in the appendix for summary statistics). Strikingly, all of the available metrics support the notion that the empathy video was more successful at grabbing attention, and for longer periods of time: On average, users watched more of the video, the share of people viewing the video for at least three seconds was higher, and the number of clicks on the video and on the forwarding button were higher per impression and per €1 spent. For the three variables computed as shares, we tested treatment

[24]The empathy video can be accessed here: `https://www.youtube.com/watch?v=vNIKofWG6iE` and the effectiveness video at: `https://www.youtube.com/watch?v=KFSQjLATgnU`.

[25]We refer to a two-sided test. In contrast, for Column (1), a one-sided test supports a conjecture that the empathy coefficient is larger than the effectiveness coefficient at $p<0.1$.

Table 5. Effects of Additional Treatment Variation

	Short term		Long term	
Dependent variable:	Per million inhabitants per day			
	Donation revenue	Donation frequency	Donation revenue	Donation frequency
	(1)	(2)	(3)	(4)
Panel A: Video type				
Empathy video	27.501*	0.420***	14.575	0.225***
	(15.935)	(0.147)	(11.428)	(0.085)
Effectiveness video	3.957	0.242	20.732*	0.196**
	(12.866)	(0.149)	(11.557)	(0.086)
Controls	yes	yes	yes	yes
Randomization blocks FEs	yes	yes	yes	yes
Observations	7,686	7,686	7,686	7,686
R^2	0.177	0.189	0.205	0.283
Panel B: Impression allocation strategy				
Fixed postal-code budgets	8.690	0.234	16.710	0.173**
	(13.770)	(0.143)	(11.830)	(0.084)
Free allocation	22.746	0.428***	18.590*	0.248***
	(15.075)	(0.154)	(11.148)	(0.087)
Controls	yes	yes	yes	yes
Randomization blocks FEs	yes	yes	yes	yes
Observations	7,686	7,686	7,686	7,686
R^2	0.176	0.189	0.205	0.283

Notes. See notes to Table 2.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

differences using a test of proportions. The differences were highly significant for clicks per impression and for the share of people watching the video for at least three seconds. The differences were not significant for clicks on the forwarding button per impression. For the other variables, we could not test treatment differences reliably, as they are based on semi-aggregated and not individual data. Together with the results from Table 5, we conclude that the empathy video was more effective at grabbing short-term attention but, in the long term (the combined effect), the effectiveness video performed no worse that the empathy video.[26] Importantly, this suggests that relying on clicks might be misleading when comparing the effectiveness of different campaigns. Campaigns that attract more attention may not be the ones to generate higher donations.

Table 6. Clicks and Impressions by Treatments: Intermediate Outcomes

	Video type		Impression allocation strategy	
Treatment:	Empathy	Effectiveness	Free allocation	Fixed postal-code budgets
	(1)	(2)	(3)	(4)
Number of seconds video watched[a]	4.213	3.603	4.047	3.791
Video clicks per million impressions[b]	7,182.1	6,635.8***	7,443.5	6,365***
Forwarding button clicks per million impressions[b]	705.5	649.4	692.7	664.1
Video views of at least 3 seconds per million impressions[b]	232,823.3	213,740.2***	219,409	228,428.5***
Video clicks per €100 spent	48.975	43.849	49.8	42.939
Forwarding button clicks per €100 spent	4.811	4.291	4.634	4.480

Notes. Based on semi-aggregated data. For the treatment with fixed postal-code budgets, data are available at the PLZ-day level. In the free allocation treatment, the ads were targeted to a group of postal codes—separately for the empathy and effectiveness groups. Therefore, the click data are semi-aggregated; at the daily level and for all postal codes in the respective group.[a] Data weighted by impressions at each level of disaggregation in order to arrive at the correct averages.[b] We tested treatment differences for three outcomes that could be computed as shares (video clicks per impression, forwarding button clicks per impression, and video views of at least 3 seconds per impression) using the test of proportions and mark significant differences in the Columns (2) and (4). The presented numbers are rescaled per million impressions. Associated summary statistics are presented in Table A6 in the appendix.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[26] While we regard the results in Table 5 as ultimately the best specification to assess the total effect of both treatments (that is, including later donations and donations through other channels), we lacked data on donations resulting from clicks on the forwarding button after watching the video, since tracing at the level of Save the Children did not work as intended.

4.2 Degree of control over the Facebook algorithm

The literature has documented algorithmic bias in advertising assignment on Facebook such that cheaper demographic groups have a higher probability of receiving impressions. For example, Lambrecht and Tucker (2019) found discrimination against young females. In our context, this means that allowing Facebook to distribute the available budget freely between postal codes could result in choosing less expensive individuals, possibly from postal codes with lower donation potential. We tested for differences between allowing Facebook to distribute impressions freely and distributing the budget to postal codes proportionally based on Facebook reach and our estimated donation potential. The second approach allowed us to gain more control over the distribution of impressions between postal codes.

Table 5, Panel B, presents the results of the regressions following the main specification (with controls) but now differentiating the two strategies regarding the allocation of impressions between postal codes. The coefficients on the free allocation dummy are higher than those on the fixed postal-code-budget dummy and significantly different from zero in all but one specification. Any treatment differences, however, are not significant and get smaller over time. Overall, we conclude that both approaches led to similar results and that, if there was any bias in the distribution of impressions by Facebook, it did not hurt the campaign outcome (if anything, the opposite is true).

Those results can be compared to the intermediate effectiveness indicators presented in Table 6, Columns (3) and (4). The treatment with free allocation of impressions seemed to be more effective according to all of the outcomes presented except for the share of users spending more than three seconds on the page with the video (statistically significant). In this case, intermediate and comprehensive measures mostly point in the same direction: They indicate a positive effect of granting full freedom to the Facebook algorithm in a fundraising context.[27]

[27] In line with previous regression results, the combination of the empathy video and free allocation leads to the highest donation levels and frequency in the short term (see Table A7 in the appendix with 2x2 separate coefficients) in line with the intermediate metrics, (see Table A8 in the appendix). However, long-term results do no longer favor this combination, which may potentially mislead deci-

4.3 Robustness and Discussion

4.3.1 Robustness

In the following, we discuss a number of robustness checks and present some additional analyses. First, for our main specification in Columns (2) and (4) of Table 2, we show in Figure A4 in the appendix randomization inference tests that have recently become quite common (Heß, 2017; Young, 2018; Cohen and Dupas, 2010). Fisherian randomization inference provides the means to assess whether an observed realization could be observed by chance even if the treatment had no effect. This test permutes the treatment and control status in the sample and re-estimates the coefficients using this placebo assignment multiple times (we set this to 5,000). The results of this test suggest that it is unlikely that our estimates have come about by chance. Second, in Figure A5 in the appendix, we also study the sensitivity of the coefficients to the number of days after the campaign that were included in the analysis. The graphs show 90% and 95% confidence intervals. Adding days after the fundraising campaign first reduced the coefficients in line with a weaker effect outside of the treatment period. The coefficient in the donation frequency regression remained quite stable from day 16 after the campaign. Adding more days towards the end of the year again increased the coefficient in the donation revenue regression, suggesting that the campaign generated additional higher-than-average donations toward the end of the year. In this exercise, we also used the additional 10 days of data in the new year that we had access to but did not use in the main analysis. The coefficients slowly decreased in size and precision when we added days in the new year. This reflects the tradeoff between adding more observations and the fading effects of the campaign in line with Lewis et al. (2015).

Table 7. Decomposition of the Long-term Treatment Effect into its Constituent Additive Parts

	Dependent variable: donation frequency per million inhabitants per day				
	Total effect (1)	New donors (2)	Repeat donors (3)	One-time donation (4)	Recurring donation (5)
	Panel A: Short term				
Video fundraising	0.331***	0.081*	0.250**	0.289**	0.042
	(0.125)	(0.047)	(0.115)	(0.121)	(0.029)
Controls	yes	yes	yes	yes	yes
Randomization blocks FEs	yes	yes	yes	yes	yes
Observations	7,686	7,686	7,686	7,686	7,686
R^2	0.189	0.166	0.184	0.183	0.165
	Panel B: Long term				
Video fundraising	0.211***	0.099***	0.112*	0.186***	0.024
	(0.073)	(0.038)	(0.060)	(0.070)	(0.017)
Controls	yes	yes	yes	yes	yes
Randomization blocks FEs	yes	yes	yes	yes	yes
Observations	7,686	7,686	7,686	7,686	7,686
R^2	0.283	0.187	0.285	0.283	0.175

Notes. See notes to Table 2. Column (1) shows the coefficient from Table 2, Column (4). Columns (2) through (5) decompose this coefficient into its constituent parts depending on donor type and donation frequency.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.3.2 Decomposition of the treatment effect

As shown in Table 7, we reran our preferred frequency regressions, decomposing the treatment effect by donor types and donation frequency as provided by Save the Children.[28] In Columns (2) and (3) and in Columns (4) and (5) the coefficients sum up to the total effect, shown in Column (1). In the short term, the additional donations came predominantly from repeat donors, while in the long term the share of new and repeat donors were approximately equal. The donations were mostly done in a form of a one-off donation. Further decompositions by donation source and by donation type are provided in Table A9 in the appendix. Since none of the coefficients in Table A9 are negative, we do not find any indication of channel substitution within Save the Children.

sion makers who rely on impression-related quality criteria. Facebook seems to maximize engagement with the ad, which in our case was best achieved by granting Facebook maximum freedom in combination with the empathy video. This might, however, not lead to the highest donation revenue in the long term.

[28]Note that due to the level of aggregation, we can only look at giving frequency as the outcome variable.

4.3.3 Spillovers

In terms of potential spillover effects, one type of spillover in our experiment may have arisen when Facebook made mistakes in assigning postal codes, for example, by wrongly assigning people to cities if they work and spend a lot of time there.[29] Another type of spillover could have occurred if treated individuals told people in untreated postal codes about the campaign (Alatas et al., 2016; Banerjee et al., 2019; Drago et al., 2020). In order to study this issue, we added to our main long-term specification a variable indicating a share of treated postal codes within 30 kilometers.[30] We chose 30 kilometers because only 20% of employees in Germany commuted longer distances in 2017.[31] Columns (2) and (5) in Table 8 show the results. The coefficient on treatment remains significant and the magnitude remains constant compared to our main long-term results shown in Columns (1) and (4). The effect of more postal codes within 30 kilometers being treated is positive and significant. In Columns (3) and (6), we provide separate estimates by postal-code status (rural or urban) interacted with the share of nearby urban postal codes that were treated as well as with the share of nearby rural postal codes that were treated. Here, we observe that the spillovers predominantly arose from the urban postal codes and that the rural postal codes were the ones most affected. Altogether, the results suggest the existence of spillover effects. Note that given the presence of spillovers, our main results provide lower bound estimates for the effects of the campaign. These estimates suggest a total effect—a direct effect plus spillovers—of video fundraising of €170.30[32], significant at $p<0.01$, or 1.13[33] in additional donations, significant at $p<0.01$.

[29] Faizullabhoy and Korolova (2018) tested location targeting on Facebook and confirmed that targeted households received advertising suggesting high precision.

[30] The distance calculation is based on centroids. The postal codes do not need to share a border.

[31] https://heimat.bund.de/atlas/pendlerdistanzen-und-pendlerverflechtungen/, viewed on January 24, 2020.

[32] 17.661 (Table 8, Column(2))+ 0.668 (share of treated neighbors, see Table A10, Column (2)) x 228.503 (Table 8, Column(2)).

[33] As above.

Table 8. Spillover Effects from Postal Codes up to 30 Kilometers (Long term)

Dependent variable:	Per million inhabitants per day					
	Donation revenue			Donation frequency		
	(1)	(2)	(3)	(4)	(5)	(6)
Video fundraising	17.652*	17.661*	17.340*	0.211***	0.213***	0.210***
	(9.630)	(9.600)	(9.574)	(0.073)	(0.073)	(0.072)
Share of treated neighbors[a]		228.503***			1.376**	
		(74.776)			(0.602)	
Urban x share of neighbors[a] treated and urban			345.768***			2.051***
			(95.509)			(0.708)
Rural x share of neighbors[a] treated and urban			511.179***			4.151***
			(92.806)			(0.755)
Urban x share of neighbors[a] treated and rural			96.095			0.671
			(83.901)			(0.681)
Rural x share of neighbors[a] treated and rural			214.499***			1.185**
			(74.977)			(0.599)
Controls	yes	yes	yes	yes	yes	yes
Randomization blocks FEs	yes	yes	yes	yes	yes	yes
Observations	7,686	7,673	7,673	7,686	7,673	7,673
R^2	0.205	0.207	0.211	0.283	0.284	0.290

Notes. See notes to Table 2. [a]Neighbors are defined as postal codes up to 30 kilometers (centroid to centroid) and do not need to share a border. The sample is slightly smaller than the original: The shapefile is missing for a few postal codes due to administrative changes.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.3.4 Profitability

Next, we discuss the profitability of the campaign. From a fundraiser's perspective, it is not enough to know whether online fundraising generates new giving. The fundraiser also needs to know whether revenue net of the costs is positive. To calculate immediate profits, we multiplied the estimated effect per day per million inhabitants by 52 (the treated population was 52 million) and 52 days. Based on the estimate in Table 2, Panel B, Column (2), we arrived at a total of €47,726 in additional donations in the long term. However, the confidence intervals (CI) are wide. For a 90% CI, the range is between €4,892 and €90,571. This can be contrasted with the direct costs of the campaign of €33,700, such that the direct revenue based on the point estimate was €1.45[34] per €1 spent. While it is easy to calculate an immediate net effect, this might be misleading. Some new donors are expected to become recurring donors, so

[34]90% CI of 0.15–2.74.

each donation has a multiplicative value. Assuming a lifetime value of a new donor of 1.75[35] and similar effects for existing donors,[36] we arrived at €2.53[37] in additional donations for each €1 initially spent. This long-term estimate is, however, below industry standards, which characterize fundraising costs of a maximum of 30% as acceptable.[38] Given that we ran a largely untargeted campaign our estimates can be regarded as lower bound estimates with a large level of external validity with respect to potential donors. Higher returns would be expected if charities were to run more conservative campaigns that target the most promising potential donors. We will address this in the next subsection.

The results of the campaign should also be considered in light of the available and comparable alternatives. Such alternatives include direct mailing to the general public. For a given campaign budget of €33,700, a charity could send around 80,000 letters (counting the costs of print and mailing but not of purchasing the addresses). Still, even with a return rate of a half of a percentage point[39] and an average donation of €87 as found in our context such a campaign would likely underperform compared to the results of our online campaign.

4.3.5 Heterogeneity

Next, we studied the heterogeneity of our treatment effect. We used the available characteristics of the postal codes and we binarized continuous variables to create

[35] In our data, around 30% of new donors chose the option of a recurring donation. Adena and Huck (2019) documented that 36.5% of donors in the first year donated again in the second year, and among those who donated twice, the return rate was 61%. Our review of online resources shows that numbers around 30% and 60% are commonly provided as estimates for first-year and later-on retention rates (see Table A11 in the appendix). Assuming that a discount factor is counterbalanced by increases in donation value, this leads to a lifetime value (LTV) of $1 + 0.3/(1 - 0.6) = 1.75$.

[36] The literature on charitable giving has documented substantial persistence in donation choices. Charitable giving in one year is the best predictor of giving in the following year (Meier, 2007; Landry et al., 2010), and the amounts chosen are usually very close to previous ones (Adena and Huck, 2022). Furthermore, treatment-imposed differences in gift level can still be observed in later gifts after the treatment has ceased to apply (Adena et al., 2014).

[37] 90% CI of (0.26–4.80).

[38] See 4.b.(2) on page 17 of https://www.dzi.de/wp-content/pdfs_DZI/DZI-SpS-Leitlinien_2019.pdf, viewed on April 14, 2022.

[39] Rates of 0.5 of a percentage point or less are to be expected from a fundraising letter to the general population. For example, Kamdar et al. (2015) documents a response rate of 0.34 of a percentage point for a standard letter in their control group.

below- and above-median dummy variables. Table 9 shows the results in our main long-term specification (with controls), in which we now interact our treatment status with the below- and above-median dummy. The results suggest that the performance of the fundraising campaign could have been greatly improved had the managers targeted postal codes with above-median shares of employed population, children, and Catholics, or those with below-median shares of native Germans, Protestants, couples, single parents, and below-median population and Facebook reach. Note that those characteristics are correlated with each other. Another good predictor of the success of a campaign is urban status of the postal code, which pertains to 16.5% of the postal codes in our sample. The best predictor of the campaign's success is the above-median estimated potential.

Table 9. Heterogeneous Treatment Effects

Video fundraising in postal codes with	Estimated potential	Population	Facebook reach	German nationals	Share Catholics	Share Protestants	Share employed	Share couples	Share children	Share single parents	Share green party voters		Urban status
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)		(12)
Panel A: Donation revenue per million inhabitants per day													
Below median	-32.047***	49.628***	55.897***	50.536***	-10.344	21.888*	-10.498	34.839***	6.259	46.255***	-9.287	Urban	47.807**
	(10.641)	(14.225)	(14.019)	(13.611)	(11.329)	(12.015)	(10.658)	(12.769)	(11.788)	(12.622)	(11.793)		(19.719)
Above median	63.735***	-16.627	-20.446*	-18.195	43.361***	11.166	43.341***	-1.287	26.464**	-13.583	41.746***	Rural	10.143
	(13.218)	(10.617)	(11.034)	(11.425)	(12.604)	(11.506)	(13.265)	(11.399)	(12.232)	(10.871)			(10.041)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes		yes
Randomization blocks FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes		yes
Observations	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686		7,686
R^2	0.195	0.192	0.192	0.193	0.192	0.190	0.192	0.191	0.190	0.192	0.191		0.190
Panel B: Donation frequency per million inhabitants per day													
Below median	-0.287***	0.531***	0.641***	0.437***	-0.100	0.317***	0.061	0.340***	0.038	0.400***	-0.079	Urban	0.393***
	(0.085)	(0.109)	(0.117)	(0.089)	(0.083)	(0.091)	(0.090)	(0.091)	(0.084)	(0.093)	(0.087)		(0.116)
Above median	0.673***	-0.134*	-0.216***	-0.045	0.498***	0.081	0.335***	0.061	0.355***	-0.006	0.470***	Rural	0.159**
	(0.097)	(0.076)	(0.077)	(0.094)	(0.094)	(0.086)	(0.090)	(0.090)	(0.093)	(0.084)	(0.092)		(0.079)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes		yes
Randomization blocks FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes		yes
Observations	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686		7,686
R^2	0.273	0.267	0.269	0.266	0.268	0.264	0.264	0.264	0.265	0.266	0.267		0.264

Notes. See notes to Table 2. Control variables contain only the lagged dependent variable.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Conclusions

This paper has explored whether online fundraising can prompt charitable giving. By randomly assigning Save the Children fundraising videos on Facebook to almost all of Germany's 8,181 postal codes, we found that an online fundraising campaign significantly increased total donations to Save the Children. Reassuringly, the largely untargeted campaign was profitable for the fundraiser: €1 spent translated into an immediate return of €1.45 and is expected to turn into €2.53 in the long run.[40] This shows that the "power of asking" (Yörük, 2009; Andreoni and Rao, 2011) also works in an online context, in which "social pressure" is clearly lower (Adena and Huck, 2020). However, we also detected some substitution between charities and projects in response to the Save the Children fundraising campaign. This suggests that fundraising might not expand individuals' donation budgets (Thaler, 1985) and that the money spent on fundraising could merely cause some redistribution and thus be ultimately lost to the charitable sector.

Our design advances the growing literature on online fundraising and advertising in several key ways. First, we use a geo-randomized experiment across all of Germany. Doing so ensured that our results have a high degree of external validity while achieving reasonable statistical power. Second, by analyzing all of the donations made to the charity, we captured the total effect of the campaign, ensuring that our results are not biased by potential substitution across channels and intertemporal substitution by donors. Third, our design addresses the question of substitution between charities and the question of individual donation budgets. Fourth, by analyzing donation data over a period of 12 weeks, we covered an extended time period and can speak to the long-term effects of online fundraising, which are more promising than previously believed. Fifth, by comparing results based on intermediate metrics like click-through rates and time spent watching videos with results based on total donations, we showed that such intermediate metrics might be misleading. This is of great importance for professional fundraiser and advertiser, charities and firms, and academic researchers, who often rely on intermediate metrics when evaluating campaigns, although the ultimate relevant outcome is (donation) revenue.

Based on our results, we see three fruitful avenues for future research. First, to uncover the mechanisms, we randomized whether the videos highlighted empathy or

[40]The numbers are based on our point estimates and the assumption that the LTV is 1.75.

the charity's effectiveness. While the empathy video was more successful in the short term, in the long term, the differences between the treatments were not significant. The modest differences suggest that the mechanism increasing charitable giving is simply the donation ask. Future studies could help to determine whether a mere impression of the charity and a subsequent call-to-action to donate is sufficient. Put differently, long videos may not be necessary to increase charitable giving.

Second, we also randomized whether Facebook's algorithm was allowed to distribute ads freely or whether we specifically allocated budgets to postal codes proportional to size and donation potential. The seemingly better performance of the free-allocation treatment calls into question the hypothesized negative effect of the Facebook algorithm, at least for charities. If the algorithm optimizes engagement—one plausible conjecture—this likely helps charities that are trying to generate new giving. The situation may, however, be quite different for other advertisers. If a luxury car manufacturer sees its ads sent to postal codes with high engagement, it is possible that the individuals in those postal codes will not be potential customers.

Third, our experiment did not test individual-level targeting, that is, any given resident in a postal code (subject to Facebook's algorithmic assignment) received the same video. Future studies could explore whether sending empathy videos to those individuals most likely to react to such content is a more effective strategy. While this comes at the cost of drawing causal inferences for the general population, it may help charities boost charitable giving more effectively. After all, the fact that a largely untargeted campaign increased donations by meaningful amounts indicates that online advertising is a highly effective fundraising tool. The relevance of our findings is clear given that online activities will likely continue to grow in importance for the nonprofit sector.

References

Adena, M. (2016). Nonprofit organizations, free media and donor's trust. *Journal of Economics*, 118(3):239–263.

Adena, M. and Huck, S. (2019). Giving once, giving twice: A two-period field experiment on intertemporal crowding in charitable giving. *Journal of Public Economics*, 172:127–134.

Adena, M. and Huck, S. (2020). Online Fundraising, Self-Image, and the Long-Term Impact of Ask Avoidance. *Management Science*, 66(2):722–743.

Adena, M. and Huck, S. (2022). Personalized fundraising: A field experiment on threshold matching of donations. *Journal of Economic Behavior & Organization*, 200:1–20.

Adena, M., Huck, S., and Rasul, I. (2014). Charitable Giving and Nonbinding Contribution-Level Suggestions Evidence from a Field Experiment. *Review of Behavioral Economics*, 1(3):275–293.

Alatas, V., Banerjee, A., Chandrasekhar, A. G., Hanna, R., and Olken, B. A. (2016). Network structure and the aggregation of information:theory and evidence from Indonesia. *American Economic Review*, 106(7):1663–1704.

Altmann, S., Falk, A., Heidhues, P., Jayaraman, R., and Teirlinck, M. (2018). Defaults and Donations: Evidence from a Field Experiment. *Review of Economics and Statistics*, 101(5):808–826.

Andreoni, J. (1995). Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments. *The Quarterly Journal of Economics*, 110(1):1–21.

Andreoni, J. and Rao, J. M. (2011). The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of Public Economics*, 95(7-8):513–520.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., and Walton, M. (2017a). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, 31(4):73–102.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2019). Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials. *Review of Economic Studies*, 86(6):2453–2490.

Banerjee, A., Chassang, S., and Snowberg, E. (2017b). Decision Theoretic Approaches to Experiment Design and External Validity. In Banerjee, A. V. and Duflo, E., editors, *Handbook of Economic Field Experiments (Vol. 1, pp. 141-174)*. North-Holland.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., and Zinman, J. (2010). What's advertising content worth? Evidence from a consumer credit marketing field experiment. *The Quarterly Journal of Economics*, 125(1):263–306.

Bilodeaua, M. and Slivinski, A. (1997). Rival charities. *Journal of Public Economics*, 66(3):449–467.

Blake, T., Nosko, C., and Tadelis, S. (2015). Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment. *Econometrica*, 83(1):155–174.

Bøg, M., Harmgart, H., Huck, S., and Jeffers, A. M. (2012). Fundraising on the internet. *Kyklos*, 65(1):18–30.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298.

Castillo, M., Petrie, R., and Wardell, C. (2014). Fundraising through online social networks: A field experiment on peer-to-peer solicitation. *Journal of Public Economics*, 114:29–35.

Chen, Y., Li, X., and MacKie-Mason, J. (2005). Online Fund-raising Mechanisms: A Field Experiment. *Contributions to Economic Analysis & Policy*, 5(2):Article 4.

Cohen, J. and Dupas, P. (2010). Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment. *The Quarterly Journal of Economics*, 125(1):1–45.

Deryugina, T. and Marx, B. M. (2021). Is the Supply of Charitable Donations Fixed? Evidence from Tornadoes. *American Economic Review: Insights*, 3(3):383–98.

Donkers, B., van Diepen, M., and Franses, P. H. (2017). Do charities get more when they ask more often? Evidence from a unique field experiment. *Journal of Behavioral and Experimental Economics*, 66:58–65.

Drago, F., Mengel, F., and Traxler, C. (2020). Compliance Behavior in Networks: Evidence from a Field Experiment. *American Economic Journal: Applied Economics*, 12(2):96–133.

Eckel, C., Grossman, P. J., and Milano, A. (2007). Is more information always better? An experimental study of charitable giving and Hurricane Katrina. *Southern Economic Journal*, 74(2):388–411.

Faizullabhoy, I. and Korolova, A. (2018). Facebook's advertising platform: New attack vectors and the need for interventions.

Filiz-Ozbay, E. and Uler, N. (2019). Demand for giving to multiple charities: An experimental study. *Journal of the European Economic Association*, 17(3):725–753.

Gallier, C., Goeschl, T., Kesternich, M., Lohse, J., Reif, C., and Römer, D. (2023). Inter-charity competition under spatial differentiation: Sorting, crowding, and spillovers. *Journal of Economic Behavior & Organization*, 216:457–468.

Gee, L. K. and Meer, J. (2019). The Altruism Budget: Measuring and Encouraging Charitable Giving. In Powell, W. W. and Bromley, P., editors, *The Nonprofit Sector A Research Handbook, Third Edition (pp. 558-565)*. Stanford University Press.

Gneezy, U., Keenan, E. A., and Gneezy, A. (2014). Avoiding overhead aversion in charity. *Science*, 346(6209):632–5.

Grieder, M. and Schmitz, J. (2020). Moral licensing or substitution? The impact of multiple opportunities to give on contributions to charity – Evidence from the field and the lab.

Hager, A. (2019). Do online ads influence vote choice? *Political Communication*, 36(3):376–393.

Heß, S. (2017). Randomization inference with Stata: A guide and software. *Stata Journal*, 17(3):630–651(22).

Jayaraman, R., Kaiser, M., and Teirlinck, M. (2023). Charitable donations to natural disasters: evidence from an online platform. *Oxford Economic Papers*, 75(4):902–922.

Johnson, G. A., Lewis, R. A., and Nubbemeyer, E. I. (2017). Ghost Ads: Improving the Economics of Measuring Online Ad Effectiveness. *Journal of Marketing Research*, 54(6):867–884.

Kamdar, A., Levitt, S. D., List, J. A., Mullaney, B., and Syverson, C. (2015). Once and Done: Leveraging Behavioral Economics to Increase Charitable Contributions.

Kessler, J. B. and Milkman, K. L. (2018). Identity in Charitable Giving. *Management Science*, 64(2):845–859.

Krieg, J. and Samek, A. (2017). When charities compete: A laboratory experiment with simultaneous public goods. *Journal of Behavioral and Experimental Economics*, 66:40–57.

Lacetera, N., Macis, M., and Slonim, R. (2012). Will there be blood? Incentives and substitution effects in pro-social behavior. *American Economic Journal: Economic Policy*, 4(1):186–223.

Lambrecht, A. and Tucker, C. (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science*, 65(7):2966–2981.

Landry, C. E., Lange, A., List, J. A., Price, M. K., and Rupp, N. G. (2006). Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment. *The Quarterly Journal of Economics*, 121(2):747–782.

Landry, C. E., 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(3):958–83.

Lange, A. and Stocking, A. (2012). The Complementarities of Competition in Charitable Fundraising. Congressional Budget Office Washington, DC Working Paper, 32.

Lewis, R. A. and Rao, J. M. (2015). The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics*, 130(4):1941–1973.

Lewis, R. A., Rao, J. M., and Reiley, D. H. (2015). Measuring the Effects of Advertising: the digital frontier (pp. 191-218). In Goldfarb, A., Greenstein, S. M., and Tucker, C. E., editors, *Economic Analysis of the Digital Economy*. University of Chicago Press.

Lewis, R. A. and Reiley, D. H. (2014). Online ads and offline sales: Measuring the effect of retail advertising via a controlled experiment on Yahoo! *Quantitative Marketing and Economics*, 12(3):235–266.

Meer, J. (2014). Effects of the price of charitable giving: Evidence from an online crowdfunding platform. *Journal of Economic Behavior & Organization*, 103:113–124.

Meer, J. (2017). Does fundraising create new giving? *Journal of Public Economics*, 145:82–93.

Meier, S. (2007). Do Subsidies Increase Charitable Giving in the Long Run? Matching Donations in a Field Experiment. *Journal of the European Economic Association*, 5(6):1203–1222.

Müller, S. and Rau, H. A. (2019). Too cold for warm glow? Christmas-season effects in charitable giving. *PLoS ONE*, 14(5):1–13.

Petrova, M., Perez-Truglia, R., Simonov, A., and Yildirim, P. (2024). Are Political and Charitable Giving Substitutes? Evidence from the United States. *Management Science, forthcoming.*

Reinstein, D. (2011). Does One Charitable Contribution Come at the Expense of Another? *The B.E. Journal of Economic Analysis & Policy*, 11(1):Article 40.

Reinstein, D. and Riener, G. (2012). Substitution Among Charitable Contributions. Convergent Lab and Field Evidence.

Rose-Ackerman, S. (1982). Charitable Giving and "Excessive" Fundraising. *The Quarterly Journal of Economics*, 97(2):193–212.

Scharf, K., Smith, S., and Ottoni-Wilhelm, M. (2022). Lift and Shift: The Effect of Fundraising Interventions in Charity Space and Time. *American Economic Journal: Economic Policy.*

Thaler, R. (1985). Mental Accounting and Consumer Choice. *Marketing Science*, 4(3):199–214.

Yörük, B. K. (2009). How responsive are charitable donors to requests to give? *Journal of Public Economics*, 93(9-10):1111–1117.

Young, A. (2018). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. *The Quarterly Journal of Economics*, 134(2):557–598.

Online Appendix can be accessed here:

https://bibliothek.wzb.eu/pdf/2024/ii20-302r2_appendix.pdf

Discussion Papers of the Research Area Markets and Choice 2020

Research Unit: **Economics of Change**