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 2020–302r

February 2020 (revised January 2022)

Research Area
Markets and Choice

Research Unit
Economics of Change

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

Copyright remains with the authors.

Discussion papers of the WZB serve to disseminate the research results of work in progress prior to publication to encourage the exchange of ideas and academic debate. Inclusion of a paper in the discussion paper series does not constitute publication and should not limit publication in any other venue. The discussion papers published by the WZB represent the views of the respective author(s) and not of the institute as a whole.

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 volume and frequency of donations to Save the Children and other charities by postal code. Our design circumvents many shortcomings inherent in studies based on click-through data, especially substitution and measurement issues. We found that (i) video fundraising increased donation frequency and value 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 non-negligible crowding out of donations to other similar charities or projects. Finally, we demonstrated that click data are 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 2019 BBE Workshop, the 2019 BSE Micro Workshop, 2019 Groningen Cause Marketing Work shop, 2019 DICE Research Seminar, and the 2019 CESifo Workshop on Economics of Digitalization for their helpful suggestions and comments. We are grateful to Julian Harke, Katharina Dorn, Meret Borchmann, Max Padubrin, and Steffen Mayer 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).

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 for the last ten 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 treatment. Information on the nonprofit market predominantly consists of anecdotal evidence, fundraisers' intuition, and advice from for-profit consultancies (Landry et al., 2006, 2010). Yet the effectiveness criteria for advertising and fundraising differ, and 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 shortcomings. First, they are plagued by very low statistical power because donations are infrequent and volatile (Lewis and Rao, 2015).[5] Second, when donors give via a link embedded in the ad, they may simply be

[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 yearly as suggested by the Blackbaud Institute, see footnote 1.

[3] See, for example, Lewis et al. (2015) and the references cited herein. For studies on online advertisement effectiveness in the context of voting, see Bond et al. (2012) and Hager (2018, 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 later arriving donors reacted to earlier donations. On the platform DonorsChoose, Meer (2017) studied how matching grants to certain projects affected giving to other. Scharf et al. (2017) 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.

substituting away from other donation channels (Blake et al., 2015) or from giving at some other time (Adena and Huck, 2019a). Third, the opposite is also possible: Online ads may lead ad recipients to give 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. The design thus circumvents the aforementioned shortcomings in that we study changes in the overall volume of donations by postal code across all possible donation channels, thus bypassing 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 ensures statistical power and a reasonable degree of external validity. 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 of potential effects on competing charities, of spillovers, and of increasing the scope of the campaign (Banerjee et al., 2017a,b).

The results show that the largely untargeted fundraising campaign increased total donation volume and donation frequency to Save the Children during and at least in the five weeks after the campaign. The increase in donation volume is estimated to be €27.2[6] per million inhabitants per day from the average of €129.3 in the control group, while the increase in frequency is estimated to be 0.16 donations per million inhabitants per day from the control-group average of 1.80.[7] Those point estimates

[6] In our preferred long-term specification with postal code fixed effects and period fixed effects. The coefficients are significant at $p < 0.1$.

[7] As above.

translate into €2.18 in additional donations for each €1 spent in immediate returns. Assuming a realistic long-term multiplier for a new donation of 1.75,[8] this implies a return of €3.82 in the 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 a non-negligible 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., 2019). 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., 2019) rather than acting as complements (Krieg and Samek, 2017; Lange and Stocking, 2012; Filiz-Ozbay and Uler, 2019).

In order to parse 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 effectiveness video generated higher donations than the empathy video but the differences are not

[8] See Section 4.3 for details.

significant. Compared to the fixed PLZ-level budgets, the treatment that allowed Facebook to distribute impressions freely between postal codes led to higher donation frequency, but again the differences are not significant. While any conclusions are necessarily limited to the specific implementation of our campaign, we interpret these results on the additional treatment variation as adding 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 contradict results based on directly relevant measures. We see this as a clear argument against using such metrics as proxies for the behavior of interest, like donating. We therefore advise advertising and fundraising managers to use careful experimental designs that can account for substitution and long-term effects while studying relevant outcomes.

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 distinguish between two types of videos and 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 (PLZ), all of which can be targeted via Facebook's advertising manager.[9] For each postal code we knew Facebook's estimated reach, that is, the number of individuals Facebook estimates it can target. We excluded the lower 5% percent as well as the upper 1% of the reach variable for several reasons. First, since a 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.[10] 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.[11] 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 is a low frequency behavior, and we were unlikely to observe many individuals giving in both pretreatment and posttreatment periods, consequently reducing the power of our experiment to that of a simple randomized experiment. (iv) Keeping the control and treatment groups comparable in individual-level experiments requires posting unrelated ads for the control group, which is costly. (Without unrelated ads, more active individuals are more likely to receive an ad, that is, to end up in the treatment group,

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

[10] 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.

[11] This depends on the quality of the Facebook algorithm assigning postal codes to individuals. Mistakes seem more likely in city centers where people work but do not live.

but also more likely to be active in all online contexts, including online giving (activity bias, see Johnson et al., 2017).

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[12] and other salient pretreatment 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.[13] 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).[14]

[12] 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.

[13] 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 group, which is well within the margin of statistical error.

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

The natural field experiment was implemented between November 10[15] 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 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).[16] 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,[17] 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.[18] 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 make their decision to donate in December might still remember the Save the Children ad and direct their donations to that charity.[19]

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

was added.

[15]In the evening hours.

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

[17]In 2017 in Germany, the total donation volume 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).

[18]The decision to exclude the period in the new year is clearly in line with 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.

[19]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.

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 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 (PLZ). 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 1500 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.[20] For each postal code, we aggregated donation volume and the number of donations at the period level: before, during, and after the

[20]This is standard in the literature (see, for example, Kessler and Milkman, 2018). Unfortunately, we cannot winsorize at the individual level. Yet in 96% of positive PLZ-day donations only one person donated and there were only a few cases in which three or more 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.

treatment (or during together with after the treatment). Finally, we normalized those variables by population size and period length so that they measure donation value and frequency per million inhabitants per day.

3.1 Main effects

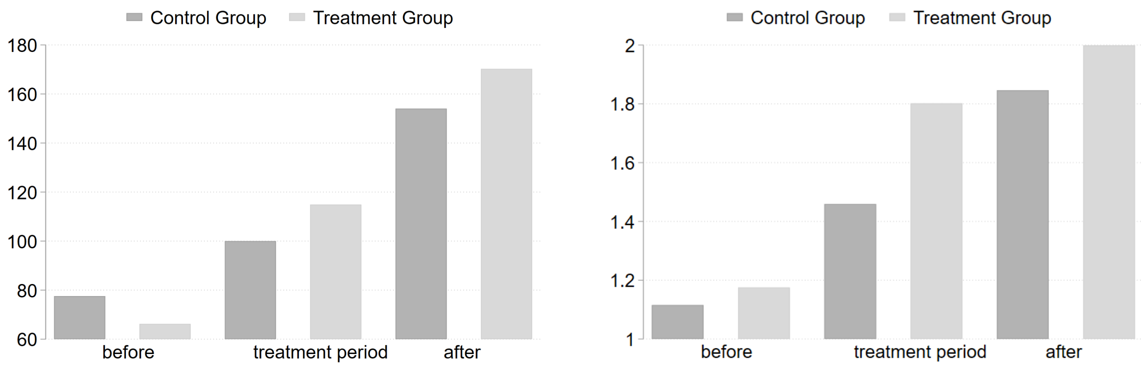
Postal codes are our unit of observation, while the outcome variables are donation value and donation frequency per million inhabitants per day. Figure 1 provides an initial overview of the data. The left panel shows the average donation amount per million inhabitants per day for three distinct periods—before, during, and after the treatment—and 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, and in addition to the Table A1 testing the balance of postal-code characteristics by treatment, we tested for the existence of pretreatment differences between the treated and untreated postal codes in a series of regressions. Table A3 in the appendix presents the results. We can confirm that there were no statistically significant pretreatment differences between the treatment and the control group in donation level or frequency.

In order to estimate the causal effect of the video fundraising campaign on donation volume and frequency, we studied both the immediate effect during the two weeks of the campaign (short term) and the more comprehensive effect, being the combined effect of during and after the campaign until the end of the year (long term). We started

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

Donation volume	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.

Table 1. Effects of Video Fundraising on Donation Level and Frequency

	Panel A: Effect during the campaign (short term)							
Dependent variable:	Per million inhabitants per day							
	Donation value				Donation frequency			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Video fundraising	14.966	37.409***	48.644***	26.228	0.344***	0.570***	0.626***	0.284**
	(11.982)	(9.136)	(9.124)	(18.083)	(0.125)	(0.089)	(0.087)	(0.135)
Postal code FE			Yes	Yes			Yes	Yes
Period FE				Yes				Yes
Postal codes	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686
Periods	T	B+T	B+T	B+T	T	B+T	B+T	B+T
Observations	7,686	15,372	15,372	15,372	7,686	15,372	15,372	15,372
R^2	0.000	0.001	0.507	0.507	0.001	0.004	0.555	0.555
	Panel B: Effect during and after the campaign (long term)							
Dependent variable:	Per million inhabitants per day							
	Donation value				Donation frequency			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Video fundraising	15.970*	67.976***	89.065***	27.232*	0.219***	0.704***	0.846***	0.159*
	(9.678)	(7.781)	(7.729)	(16.461)	(0.077)	(0.057)	(0.054)	(0.086)
Postal code FE			Yes	Yes			Yes	Yes
Period FE				Yes				Yes
Postal codes	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686
Periods	TP	B+TP	B+TP	B+TP	TP	B+TP	B+TP	B+TP
Observations	7,686	15,372	15,372	15,372	7,686	15,372	15,372	15,372
R^2	0.000	0.005	0.524	0.526	0.001	0.011	0.656	0.660

Notes. Linear estimations in Stata. Specifications (4) and (8) were run with the command areg, absorbing postal-code fixed effects. FE: fixed effects. T: treatment period. B: pretreatment period. TP: treatment and posttreatment periods pooled. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

with a simple comparison between the treatment and control groups in a cross-sectional framework. The results are presented in Table 1, Column (1) for donation value and in Column (5) for donation frequency. Panel A presents the short-term effect (treatment period only) and Panel B the long-term effect (treatment and posttreatment periods pooled). While the short term effect of video fundraising on donation value is not significant, the long term effect is significant at $p<0.1$. Both coefficients are similar in magnitude and suggest an additional €15 in donations per million inhabitants per day. Both coefficients in the regressions with donation frequency as the outcome variable are highly significant and suggest an additional 0.34 donations per million inhabitants per day in the short term and 0.22 donations in the long term. These results give us estimates from the simple randomized experiment that rest on the assumption that the particular instance of the randomization procedure achieved a good balance. Although we have no reason to doubt that this is the case, the number of units in the experiment is finite and the heterogeneity between our units of observation is large. Therefore, the estimated effects depend to a large extent on the actual allocation to the treatment and control groups. Had they been available, we could have improved the precision of the estimates by including many control variables. Unfortunately, we do not have sufficient information on variables that were potentially important but not observable, for example, average empathy in a postal code (empathy has oftentimes been shown to correlate strongly with giving behavior).[21] In Figure 1, we observe some (statistically not significant) differences in donation value and frequency in the pretreatment period. We can correct for those differences and account for characteristics of the postal codes that are fixed over time by using a difference-in-difference technique. Therefore, in Columns (3) and (7) of Table 1, we include postal-code fixed effects for which we also add the pretreatment period (in Columns (2) and (6), we also show the intermediate step of including the pretreatment data without accounting for postal code fixed effects). Given the stark differences in charitable

[21]Table A4 in the appendix shows the results when we control for the available characteristics, randomization blocks fixed effects, and past donations. The differences between coefficients are small.

giving over time, we deemed it necessary to add period fixed effects as well: we do this in Columns (4) and (8). While our preferred specification is the last one—two-way fixed effects—all of the other approaches suggest positive effects.

In our preferred specifications, we find that the video fundraising campaign generated an additional €26–€27 in donation value in both the short (not significant) and the long term. Regarding the number of additional donations, the estimates are 0.28 in the short term and 0.16 in the long term. The magnitudes of the long-term experimental effect in terms of the standard deviation for both outcome variables are comparable: They amount to 0.06 and 0.05.[22] The positive effects on donation frequency and similar magnitudes suggest that online fundraising predominantly 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 shows 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.

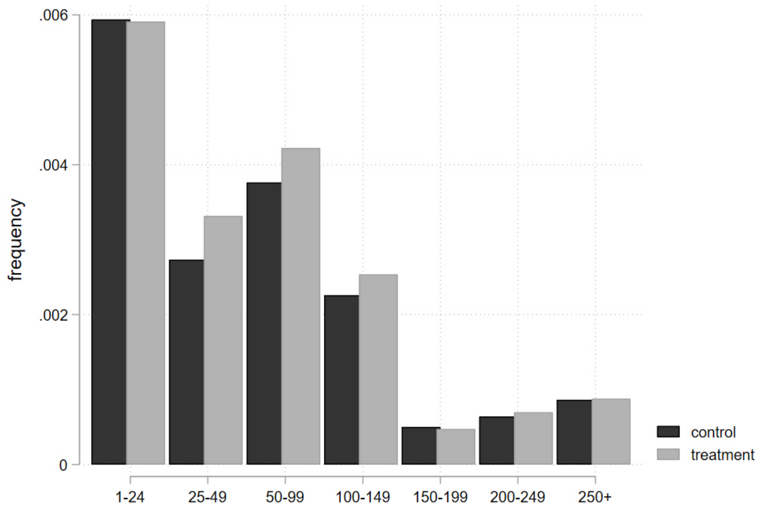
The results in Panel B of Table 1 suggest a long-term positive effect of the fundraising campaign and provide evidence countering (any sizable) intertemporal substitution. Of course, both effects might be in play and might have partly canceled each other out. Overall, we conclude that online fundraising 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 or time frame.

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

[22] Although these figures may appear small, our outcome variables are highly volatile, and we measure the effect on total donation revenue. The percentage increase in terms of the average is substantial.

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



Notes. The unit of observation is a postal code per day. In most cases, there is one donation per postal code per day. In less than 4% of observations there are two donations, and only in a few additional cases a maximum of four. Zero-donations are the omitted category.

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 volume 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 winsorize the PLZ-day level donations at €1,000. Then, for each postal code, we aggregated donation volume and the number of donations at the period level: before, during, and during together with after the treatment. Finally, we normalized those variables by population size and period length so that they measure donation value and frequency per million inhabitants per day. In Table 2, we repeat our preferred specifications from Table 1, that is, including period fixed effects and postal-code fixed effects, now using the new data. Columns (1) and (2) present the short-term results, that is, the effect during the fundraising campaign, while Columns (3) and (4) study the long-term period, which combines the campaign period and the posttreatment period. In columns (1) and (3), the dependent variable is donation value 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 giving to the other 23 charities by almost €90 per million inhabitants per day in the short term and by €60 in the long term (both significant at the 10% level). The effect on donation frequency was a (nonsignificant) reduction in the number of donations by 1.17 in the short-term and by 0.26 in the long-term.

Table 2. Effect of the Save the Children Campaign on Donations to 23 Similar Charities

	Short-term		Long-term	
	Per million inhabitants per day			
	Donation value (1)	Donation frequency (2)	Donation value (3)	Donation frequency (4)
Video fundraising	-91.101*	-1.169	-59.193*	-0.255
	(51.861)	(1.059)	(32.962)	(0.318)
Postal code FE	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes
Postal codes	7,686	7,686	7,686	7,686
Periods	B+T	B+T	B+PT	B+PT
Observations	15,372	15,372	15,372	15,372
R^2	0.552	0.510	0.602	0.571

Notes. See notes to Table 1

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

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., 2020, 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 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 is much higher than for Save the Children for the same period of time, and we observe 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 winsorize the PLZ-day-project level donations at €1,000. Then, for each postal code, we aggregate donation volume and the number of donations at the period level (and by project type): before, during, and during together with after the treatment. Finally, we normalized those variables by population size and period length so that they measure donation value and frequency per million inhabitants per day. In Table 3, we ran the same specifications as our preferred ones, that is, including period fixed effects and postal-code fixed effects. The results suggest that there was no overall effect of video fundraising on donations to other projects on betterplace.org. However, once we interact the treatment dummy with the children-related type of project, we find a negative and significant interaction effect.[23] This suggests that video fundraising by Save the Children drained donation money from other projects that benefit children.

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.[24] Our results suggest that the Save the Children fundraising campaign reduced donations to similar causes, implying a substitution effect. While there are

[23]Note that in the latter regressions, per postal code, we now have 4 observations: 2 periods of time, and, in both cases, projects related and unrelated to children.

[24]Note that in both sets of data there is an overlap and that the effects cannot be simply added: The charity alliance collects donations on betterplace.org but also via other online channels while on betterplace.org other charities are also active. The data available 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 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 displaced online fundraising on Facebook of other charities. As a consequence their ads were more likely to appear in the control group. (ii) Regular donors to other charities switched to Save the Children and chose to make lower donations to the latter.

Table 3. Effect of the Save the Children Campaign on Donations to Projects on betterplace.org

	Per million inhabitants per day							
	Short-term				Long-term			
	Donation value		Donation frequency		Donation value		Donation frequency	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Video fundraising	4.518	26.206*	-0.581	0.073	42.930	56.410***	-0.353	0.706***
	(26.233)	(15.362)	(0.468)	(0.245)	(39.102)	(19.953)	(0.411)	(0.231)
Video fundraising x children related projects		-47.494***		-0.739***		-69.450***		-1.776***
		(14.741)		(0.177)		(19.979)		(0.203)
Children related projects		-31.901***		-1.215***		-49.038***		-1.568***
		(7.668)		(0.112)		(9.594)		(0.101)
Postal code FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Postal codes	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686
Periods	B+T	B+T	B+T	B+T	B+TP	B+TP	B+TP	B+TP
Project types	-	2	-	2	-	2	-	2
Observations	15,372	30,744	15,372	30,744	15,372	30,744	15,372	30,744
R^2	0.660	0.401	0.692	0.466	0.628	0.448	0.725	0.508

Notes. See notes to Table 1.

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

papers that have studied substitution over time (Adena and Huck, 2019a) or over many charities (Meer, 2014), we were able to study both together.

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. 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 one stressing the competence and effectiveness of the organization.[25] 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 4, Columns (1) and (2), we present the results of the regressions following our preferred specification (two-way fixed effects, long-term) but now differentiate between the empathy and effectiveness videos. In Column (1) the outcome is, again, donation value per million inhabitants per day. The effectiveness video coefficient is significant and more than double the empathy video coefficient, which is insignificant. However, we cannot reject equality of the two types of videos.[26] In Column (2), which shows the results regarding the frequency of donations, the coefficients are nearly identical, although only significant for the empathy video. We conclude that there are no differences in the effects by type of video (although a difference may exist, especially for donation value, that we are underpowered to detect).

Table 4. Effects of Additional Treatment Variation (Long term)

Video type			Impression allocation strategy		
	Per million inhabitants per day			Per million inhabitants per day	
	Donation value (1)	Donation frequency (2)		Donation value (3)	Donation frequency (4)
Effectiveness video	37.710**	0.152	Free allocation	29.851*	0.213**
	(17.630)	(0.103)		(17.869)	(0.101)
Empathy video	16.754	0.167*	Fixed postal-code budgets	24.612	0.106
	(18.724)	(0.099)		(18.499)	(0.101)
Postal code FE	Yes	Yes		Yes	Yes
Period FE	Yes	Yes		Yes	Yes
Postal codes	7,868	7,868		7,868	7,868
Periods	B+PT	B+PT		B+PT	B+PT
Observations	15,372	15,372		15,372	15,372
R^2	0.526	0.660		0.526	0.660

Notes. See notes to Table 1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table 5, Columns (1) and (2), we present some intermediate metrics that point to the mechanism behind the effects of both video types (see Table A5, Columns (1) and (2) in the appendix for summary statistics). Strikingly, all of the available

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

[26] A one-sided test that the effectiveness coefficient is larger than the empathy coefficient cannot be rejected at $p < 0.1$.

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 per impression and per €1 spent were higher. For the three variables computed as shares, we tested treatment 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 4, we conclude that the empathy video was more effective at grabbing attention but that donation frequency was rather equal for both videos, and that actual donation levels might have been higher in the effectiveness treatment.[27] Importantly, we see 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.

4.2 Degree of control over the Facebook algorithm

The literature has documented algorithmic bias in advertising assignment on Facebook such that less expensive 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 in the treatment group could result in choosing less expensive postal codes with lower donation potential. We tested for differences that would arise when allowing Facebook to distribute impressions freely versus when distributing the budget to postal codes proportional to reach (which is a function of population) and

[27] While we regard the results in Table 4 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.

Table 5. Clicks and Impressions by Treatments: Intermediate Outcomes

Treatment:	Video type: Empathy	Video type: Effectiveness	Impression allocation strategy: Free allocation	Impression allocation strategy: 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. For the free allocation treatment, data are available at the daily level for the empathy and effectiveness groups separately, that is, the data are aggregated 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 A5 in the appendix.

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

our estimated donation potential. The second approach allowed us to gain more control over the distribution of impressions between postal codes.

Table 4, Columns (3) and (4), presents the results of the regressions following our preferred specification (two-way fixed effects, long-term) but now differentiate the two strategies regarding the allocation of impressions between postal codes. The coefficients on the free allocation dummy are slightly higher than the ones on fixed postal code-budget dummy and significantly different from zero in all specifications. Any treatment differences, however, are small and not significant. Overall, we conclude that both approaches lead to similar results and that, if there is 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 5, 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 point mostly in the same direction: They indicate a positive effect of granting full freedom to the Facebook algorithm in a fundraising context.[28]

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, in Table A4 in the appendix, we provide simple cross-sectional regressions that are similar to those in Columns (1) and (5) of Table 1 but include available postal code characteristics and pretreatment donations as control variables. The coefficient sizes are not very different from those presented in Columns (1) and (5) of Table 1. Second, for our preferred specification in Columns (4) and (8) of Table 1, 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. Third, 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

[28]In line with previous regression results, the combination of the effectiveness video and free allocation leads to the highest donation levels and frequency (see Table A6 in the appendix with 2x2 separate coefficients; treatment differences are not significant). The intermediate metrics, however, do not favor this combination (see Table A7 in the appendix), potentially misleading decision 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.

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 value 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).

4.3.2 Decomposition of the treatment effect

Table 6. Decomposition of the Long-term Treatment Effect into its Constituent Additive Parts

	Total effect (1)	New donors (2)	Repeat donors (3)	One-time donation (4)	Regular donation (5)
Video fundraising	0.159*	0.086*	0.073	0.140*	0.020
	(0.086)	(0.048)	(0.069)	(0.081)	(0.026)
Postal code FE	Yes	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes	Yes
Postal codes	7,868	7,868	7,868	7,868	7,868
Periods	B+TP	B+TP	B+TP	B+TP	B+TP
Observations	15,372	15,372	15,372	15,372	15,372
R^2	0.660	0.518	0.694	0.665	0.506

Notes. See notes to Table 1. Column (1) shows the coefficient from Table 1, Panel B, Column (8). Columns (2) through (5) decompose this coefficient into its constituent parts depending on donor type and donation frequency.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As shown in Table 6, we reran our preferred frequency regressions in the long term, decomposing the treatment effect by donor types and donation frequency as provided by Save the Children.[29] In Columns (2) and (3) and in Columns (4) and (5) the coefficients sum up to the total effect of 0.159, shown in Column (1). The effect seems to be mostly driven by new donors and one-off donations, suggesting that the campaign helped to broaden the charity's base rather than speaking to established customers who would have donated regardless. Further decompositions by donation

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

source and by donation type are provided in Table A8 in the appendix. Since none of the coefficients in Table A8 are negative,[30] we do not find any indication of channel substitution within Save the Children.

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.[31] 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.[32] We chose 30 kilometers because only 20% of employees in Germany commuted longer distances in 2017.[33] Columns (2) and (5) in Table 7 show the results. The coefficient on treatment remains significant and the magnitude 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 see that the coefficients on the share of nearby urban treated postal codes are positive and highly significant. 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

[30]Except the coefficient on TV, but it is small and not significant.

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

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

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

video fundraising of €140.50[34] significant at $p < 0.05$ or 1.14[35] in additional donations significant at $p < 0.01$. Table A10 in the appendix shows a version of Table 7 based on bordering postal codes. The results are less precise.

Table 7. Spillover Effects from Postal Codes up to 30 Kilometers

	Per million inhabitants per day					
	Donation value			Donation frequency		
	(1)	(2)	(3)	(4)	(5)	(6)
Video fundraising	27.301*	27.593*	28.053*	0.160*	0.162*	0.162*
	(16.487)	(16.496)	(16.531)	(0.086)	(0.086)	(0.086)
Share of neighbors[a] treated and urban		169.132**			1.470**	
		(78.672)			(0.601)	
Urban x share of neighbors[a] treated and urban			397.055***			2.570***
			(95.363)			(0.678)
Rural x share of neighbors[a] treated and urban			217.198**			2.506***
			(104.352)			(0.744)
Urban x share of neighbors[a] treated and rural			221.932***			1.764***
			(84.935)			(0.627)
Rural x share of neighbors[a] treated and rural			157.120**			1.286**
			(77.447)			(0.596)
Postal code FE	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Postal codes	7,673	7,673	7,673	7,673	7,673	7,673
Periods	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT
Observations	15,346	15,346	15,346	15,346	15,346	15,346
R^2	0.526	0.526	0.527	0.660	0.660	0.661

Notes. see notes to Table 1. [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. In order to calculate immediate profits, we needed to multiply the estimated daily effect by 52 days and 52 (the treated population was 52 million). Based on the estimate in Table 1, Panel B,

[34]27.593 (Table 7, Column(2))+ 0.668 (share of treated neighbors, see Table A9, Column (2)) x 169.132 (Table 7, Column(2)).
[35]As above.

Column (4), we arrived at a total of almost €73,500 in additional donations. However, the confidence intervals are wide. For a 90% confidence interval, the range is between €163 and €146,746. 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 €2.18 per €1 spent. While it is easy to calculate an immediate net effect, this might be misleading. Some new donors are expected to turn into regular donors, so each donation has a multiplicative value. In our sample, almost one third of new donors chose the option to donate regularly. Of course, we do not know when a donor will cancel and whether nonregular donors will give repeatedly as well. Assuming a realistic lifetime value of a new donor of 1.75[36] and similar effects for existing donors,[37] we arrived at €3.82 in additional donations for each €1 initially spent. This long-term estimate is in line with industry standards that characterize fundraising costs of a maximum of 30% as acceptable.[38] Still, lower spending on fundraising would be recommended. 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 next.

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

[36]In our data, around 30% of new donors chose the option of a regular donation. Adena and Huck (2019a) 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 A12 in the appendix). Assuming that a discount factor is counterbalanced by increases in donation value, this leads to a lifetime value (LTV) of $1 + 0.3/(1 - 0.6) = 1.75$.

[37]The literature on charitable giving has documented sizable 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, 2019b). 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).

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

[39]Rates of 0.5 of a percentage point or less are to be expected from a fundraising letter to the general

€120,[40] such a campaign would likely underperform 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 below- and above-median dummy variables. Table 8 shows the results in our preferred long-term specification, 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 populations, Facebook reach, GDP, shares of Catholics, and shares of Green Party voters in the 2009 European Union elections, or those with below-median shares of native Germans, Protestants, couples, children, and single parents. Note that those characteristics are correlated with each other. The variable estimated potential is based on a mix of those other characteristics but performs (only) slightly better than some of the single variables. An even better predictor of the success of a campaign is city status, which pertains to 16.5% of the postal codes in our sample.

4.3.6 Scaling up

Since the share of the population that received impressions varied in the treatment group, we can potentially assess the effects of scaling up such a campaign. The ratio of total impressions to the population ranges from 0 to 0.57 in the treated postal codes and can be interpreted approximately as the share of the population receiving the ad, since the majority of impressions were distributed to different individuals.[41] However, simply regressing the outcome variables on impressions per population would result in biased estimates because the ad budgets were assigned proportional to reach multi-

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.

[40]We consider an average donation of €120 to be the upper bar. This would be higher than the average donation in the current sample of €87, and it is the same average documented in Adena et al. (2014) in a relatively rich and highly educated sample of Munich opera goers.

[41]Note that we had to exclude data from the free allocation treatment since no impression data at the postal code level are available for this group.

Table 8. Heterogeneous Treatment Effects

Video fundraising in postal codes with	Population	Facebook reach	Estimated potential	GDP	German nationals	Share of Catholics	Share of Protestants	Share employed	Share couples	Share children	Share single parents	Share voters for the green party		City status
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)		(13)
Panel A: Outcome variable: Donation value per million inhabitants per day														
Below median	14.642	11.503	-16.097	0.493	63.428***	10.823	34.948*	-5.109	53.973***	36.527**	34.503*	5.826	City	108.888***
	(19.768)	(19.794)	(17.111)	(17.805)	(18.878)	(17.535)	(18.339)	(16.774)	(18.949)	(17.416)	(19.584)	(17.300)		(22.313)
Above median	39.910**	42.112**	69.955***	53.454***	-9.305	43.782**	19.641	59.547***	1.433	18.108	19.932	48.405**	Rural	10.785
	(16.424)	(16.606)	(19.137)	(18.531)	(17.424)	(18.820)	(18.036)	(19.474)	(17.430)	(18.897)	(16.658)	(19.004)		(16.896)
Postal code FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
Postal codes	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686		7,686
Periods	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT		B+PT
Observations	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372		15,372
R^2	0.526	0.526	0.527	0.526	0.527	0.526	0.526	0.526	0.526	0.526	0.526	0.526		0.527
Panel B: Outcome variable: Donation frequency per million inhabitants per day														
Below median	0.118	0.130	-0.222**	0.040	0.427***	-0.084	0.297***	-0.015	0.339***	0.147	0.268**	-0.079	City	0.546***
	(0.116)	(0.118)	(0.091)	(0.105)	(0.098)	(0.093)	(0.104)	(0.099)	(0.102)	(0.091)	(0.107)	(0.095)		(0.119)
Above median	0.201**	0.187**	0.535***	0.276***	-0.111	0.405***	0.023	0.334***	-0.015	0.171	0.050	0.394***	Rural	0.081
	(0.084)	(0.083)	(0.109)	(0.097)	(0.104)	(0.108)	(0.098)	(0.103)	(0.100)	(0.110)	(0.094)	(0.106)		(0.091)
Postal code FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
Postal codes	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686	7,686		7,686
Periods	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT	B+PT		B+PT
Observations	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372	15,372		15,372
R^2	0.660	0.660	0.662	0.660	0.661	0.661	0.660	0.661	0.661	0.660	0.660	0.661		0.661

Notes. See notes to Table 1.

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

plied by estimated potential (in the treatment group with fixed postal-code budgets). Nevertheless, the actual number of impressions in a postal code varied somewhat relative to the intended assignment due to a variety of factors such as rounding, the impossibility of spending less than €1, and daily fluctuation in prices per impression. Therefore, there was some degree of variation in the actual treatment intensity relative to the one assigned that we could exploit. Figure A6 in the appendix shows the difference between the budget assignment function and the actual number of impressions in the group with fixed postal-code budgets. We see that many dots lie very close to the 45-degree line but that there is also sizable variation, and many postal codes (with low reach and potential) did not receive any impressions.

In Table A13 in the appendix, we repeated our preferred estimation from Table 1 but exchanged the treatment dummy with actual treatment intensity relative to the one assigned. More specifically, the explanatory variable is defined as the total number of impressions in a postal code divided by the estimated potential multiplied by Facebook reach, that is, the budget assignment function. If the treatment had been realized as hypothetically foreseen, the variable would have taken values of one for all postal codes in the treatment group. While the coefficients in Table 1 give us intention-to-treat estimates, the coefficients in Table A13 can be more closely interpreted as average treatment effects (assuming that everyone treated consumed the ad). In Columns (2) and (4), we added the relative intensity squared.[42] The estimates are higher than in our main Table 1 and are more precise, since they reflect the actual implementation intensity. We see that a fully implemented campaign is expected to generate an average donation of more than €47 and an average additional frequency of more than 0.2 per million inhabitants per day. Note that the average implementation level in the treatment group is 0.47, hence implying an effect of the current treatment (intensity) of 22.26 in terms of donation volume and 0.1 in terms of frequency. This is similar to the estimates from Table 4, row 2, Columns (3) and (4). However, we also see a significant and sizable square term, suggesting decreasing returns to scale. This suggests that the optimal scale of the campaign is achieved at around 0.53–0.58 of the estimated donation potential times Facebook reach and that the returns to fundraising decrease after this threshold. This might be related to the Facebook's ad allocation algorithm which showed the video to the most active Facebook users first. When the intensity of the campaign increased, less active individuals received the ad as well. However, they might have been less likely to donate. This could also explain a potentially better performance of our free allocation treatment, which gave Facebook more flexibility in the distribution of the ads.

5 Conclusions

This paper has explored whether online fundraising can prompt charitable giving. Randomly assigning Facebook fundraising videos from the charity Save the Children across almost all of Germany's 8,181 postal codes, we found that an online fundraising

[42] Higher-order polynomials yield insignificant coefficients.

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 €2.18 and is expected to turn into €3.82 in the long run.[43] 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 non-negligible 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 questions 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 clearly 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 parse the mechanisms, we randomized whether the videos highlighted empathy or the char-

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

ity's effectiveness. While the effectiveness video generated slightly higher donations than the empathy video, the differences 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, perhaps long videos are not 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 fact that we found virtually no differences 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 received the same video (subject to Facebook's algorithmic assignment). 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. (2019a). 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. (2019b). Personalized fundraising: A field experiment on threshold matching of donations. WZB Discussion Paper SP II 2019–306r.

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., 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. *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. (2019). Inter-charity competition under spatial differentiation: Sorting, crowding, and spillovers. *ZEW-Centre for European Economic Research Discussion Paper*, (19-039):1–34.

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. pages 1–34.

Hager, A. (2018). Do Online Ads Sway Elections? Evidence from a Field Experiment. *Mimeo.*

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. (2020). Charitable Donations to Natural Disasters: Evidence From an Online Platform.

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. (2019). Are Political and Charitable Giving Substitutes? Evidence from the United States. SSRN Electronic Journal.

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. (2017). Lift and Shift: The Effect of Fundraising Interventions in Charity Space and Time. CESifo Working Paper Series.

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/2022/ii20-302r_appendix.pdf

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice

Discussion Papers of the Research Area Markets and Choice 2020

Research Unit: **Economics of Change**

Anselm Hager and Justin Valasek Refugees and social capital: Evidence from Northern Lebanon	SP II 2020-301
Maja Adena and Anselm Hager Does online fundraising increase charitable giving? A nationwide field experiment on Facebook	SP II 2020-302

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice