Journal of Public Economics 95 (2011) 403–415

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

The demand for products linked to public goods: Evidence from an online field experiment

Brian McManus [a,*], Richard Bennet [b]

[a] Department of Economics, University of North Carolina, Chapel Hill NC 27599-3305, United States
[b] CCA Global Partners, 4301 Earth City Expressway, St. Louis, MO 63045, United States

ARTICLE INFO

Article history:
Received 7 May 2009
Received in revised form 1 March 2010
Accepted 21 April 2010
Available online 6 May 2010

JEL classification:
C93
D64
L30
M14

Keywords:
Field experiments
Nonprofit organizations
Corporate social responsibility
E-commerce

ABSTRACT

We conduct a field experiment at a nonprofit organization's online store to study how demand changes when consumers' purchases can generate revenue for a charitable cause. When purchases can trigger a small donation by an outside anonymous group, consumers respond strongly and apparently without regard for the specific conditions that trigger the donation. Consumers respond similarly when the outside donation requires a personal donation which consumers generally decline. When the outside donations are relatively large, however, consumers appear to pay close attention to the trigger conditions, and increase their purchases only where needed to generate the outside donation. Overall, increasing the salience of financial incentives weakens consumers' positive responses to the outside group's donation pledges. We also present evidence that the donation pledges have positive long-term effects on demand and may reduce price sensitivity.

1. Introduction

An increasing number of firms and organizations are selling products linked to social causes.[1] For-profit firms do this through donation pledges (e.g. Target and Whole Foods) or commitments to green or fair trade production (e.g. Starbucks). Nonprofit organizations sell products to supplement revenue from direct donations (e.g. the Nature Conservancy offers apparel and a magazine). There may be many motivations for this activity. On the supply side, a for-profit firm may offer these products for strategic reasons, perhaps because consumers or employees value the charity attribute, while non-profits may view the products as a way to advertize their causes and expand their base of supporters. On the demand side, positive responses from consumers may come from a desire to contribute to a public good, a personal utility benefit from acting charitably, or an inference that socially-linked products are of relatively high quality. With a variety of potential motivations for their actions, consumers' observable responses to these products may be complex as well.

To improve our understanding of consumers' responses to socially-linked products, we analyze data from a field experiment conducted at the online store of a large anonymous nonprofit organization (NPO). We observe how consumers' purchasing decisions were affected by variation in the revenue their choices generated for the NPO's charitable mission. To interpret consumers' choices, which are counterintuitive in some cases, we offer a theoretical framework for describing how consumers' preferences interact with the incentives they faced within the experiment. We also study whether consumers' price sensitivity was affected by variation in the charity revenue their purchases generated, and whether the experiment's immediate effects on demand led consumers to substitute away from future purchases or direct charitable donations.

Our examination of actions both within and beyond the experimental period is new to field studies in this area. Numerous surveys and classroom experiments have established that consumers will express a willingness to favor products sold by socially responsible firms. The controlled and low-cost nature of these studies permits inference on some fine distinctions about consumers' responses, for example what types of products are most amenable to social links (Strahilevitz and Myers, 1998), whether positive or negative news has a greater impact on demand (Sen and Bhattacharya, 2001), and how consumers interpret profit donation pledges (Olsen et al., 2003). Field studies, by contrast, are more costly to conduct as experiments and more likely to contain identification challenges when they employ observational data. As a result, much of the existing literature utilizing

* Corresponding author.
E-mail addresses: mcmanusb@email.unc.edu (B. McManus), rbennet@ccaglobal.com (R. Bennet).

[1] For a description of recent trends, see Bonini et al. (2006).

0047-2727/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpubeco.2010.04.005

field data has tended to focus on testing simpler hypotheses of whether social links, in general, affect demand. Elfenbein and McManus 2010 and Popkowski Leszczyc and Rothkopf (2010) study charity auction markets and find that consumers will submit higher bids in a charity auction compared to a non-charity auction for the same product. Hiscox and Smyth (2008) use a field experiment to show that consumers will favor products with a "fair trade" label. In these field studies, it is difficult to infer what aspects of a charity announcement are most important: Is it the presence of any offer at all? The precise financial terms of the charity pledge? Some recent progress on these questions is offered by Elfenbein et al. (2010), who find that consumers view eBay sellers' charity pledges as a substitute for seller reputation. In these studies, an additional challenge lies in uncovering the broader effects of a consumer's purchase of a charity-linked product, whether through the consumer's later demand for merchandise or his donations to the same cause.

Our study also complements the substantial literature on consumers' choices in making direct charitable donations. Recent research on donation choices, largely with field experiments, demonstrates that consumers behave in interesting and sometimes surprising ways, especially with regard to varying information about others' actions. Eckel and Grossman (2003) find that financially equivalent inducements for donations can bring different results depending on how the inducements are presented, with matching pledges from other donors out-performing rebates. Karlan and List (2007) also report that matching pledges can stimulate donations significantly, but they find that substantially different matching incentives yield similar responses from donors. Shang and Croson (2009) offer further evidence that one consumer's actions can stimulate another's giving; they find that public radio donors increase their donations when told that another individual recently made a large gift. While the immediate effects of information and matching pledges can be substantial, Meier (2007a) finds that consumers who received a short-term matching stimulus reduced their later donations so that their overall contributions were no greater than those of a control group that received no stimulus. In all, these studies suggest that financial incentives matter, but the information conveyed through experimental treatments may have a greater effect. The data from our experiment suggests the same.

For our study, we observe the actions of over 100,000 consumers at the NPO's online store during March and April 2007. When each new consumer arrived at the store, a "cookie" was placed in his or her internet browser to record and preserve the random assignment of the consumer into an experimental treatment. The experiment varied the appearance of the store's front page. Consumers in a control group received the standard storefront. Others were told that $1 or $5 would be donated by an anonymous outside group if the consumer purchased at least $10 in merchandise. Yet other consumers could trigger $1, $5, or $10 donations from the outside group by purchasing at the store and donating to the NPO. All consumers received either standard store prices or discounted prices. For 12 months following the experiment, the NPO tracked consumers who purchased from the store, and we observe their subsequent store purchases and donations to the NPO.

Consumers who received donation pledges generated 20% more revenue, on average, than consumers who received no pledge.[2] Although a positive demand response to the donations pledges was expected, there are surprising aspects of the results. First, despite the structure of the incentives, the additional revenue largely came from increased order sizes. While the median consumer who purchased under the control message spent $28, the median consumer with a donation pledge spent $40. The structure of our pledge implies that this difference did nothing to increase the outside donation to the NPO. This happened even if the pledged donation was small ($1), and more strikingly if the pledged donation was not triggered at all because the consumer did not complete the required personal donation portion of an offer. We describe the donation pledges as stimulating a change in consumers' perceived value of supporting the NPO through their own purchases. Not all donation pledges had this effect. Consumers who observed the richest offer–a $5 outside donation with a purchase but no consumer donation required–had the weakest response among all consumers who received pledges. Consumers ignored these $5 pledges unless their orders risked being too small to trigger the pledged donation. These consumers appear to focus on the financial terms of the pledge rather than using the pledge to update their preferences for supporting the NPO in general. This suggests that consumers' responses to socially-linked products may be very sensitive to how a firm or nonprofit describes the public goods benefit of purchasing the products. This also provides a potential explanation for the literature's broad range of estimates for charitable giving's price elasticity, which Karlan and List (2007) describe. Small differences across studies in how price variation is generated may have a large impact on how consumers interpret the variation.

During the year following the experimental period, consumers who initially purchased under a donation pledge were more likely to return to the store and order again. This suggests that the brief stimulation of the experiment did not merely shift consumer spending away from future time periods. Instead, consumers may have interpreted the experimental pledges as positive quality signals.[3] We also find very little overlap between the store customer population and the NPO's direct donors, suggesting minimal substitution between intensified store activity and donations. Our evidence on consumers' responses to changing prices is mixed. The median consumer appears to be less price-sensitive when exposed to messages about pledged donations, but the average effects show no significant difference between consumers who received the control and a donation pledge.

The rest of the paper proceeds as follows. We introduce the experimental setting and procedures in Section 2, and in Section 3 we discuss the incentives generated by the experiment. In Section 4 we describe the data, and Section 5 contains our econometric analysis. In Section 6 we offer a concluding discussion and provide suggestions for future work.

2. Experimental setting and structure

2.1. The online store

At its online store the NPO offers a variety of books, functional equipment consistent with the NPO mission, and apparel. Product prices ranged from $4.95 to $65 at the time of the experiment. The store also includes a donation opportunity with which consumers can "buy" donations in $10 increments. In addition to the central online store at which our experiment ran, some of the NPO's products can be purchased through local branches of the organization. In total, sales at the main store account for about 15% of NPO merchandise sold through all channels. Merchandise sales are a small fraction of overall NPO annual revenue, which exceeds $500 million.

The NPO advertises the store on its home page and through emails to its past donors, customers, and volunteers. In addition, the NPO places store ads on internet search engines and in other websites'

[2] With this statistic and others referenced in the Introduction, we describe activity in the lower 95% of all store purchases. The data are skewed by a small number of large orders that were from firms rather than households. See Section 4 for discussion of this approach to the data.

[3] Landry et al. (forthcoming) report a similar effect. They observe that individuals who were stimulated to donate via an economic incentive mechanism, which could send a positive signal about a cause, were more likely to donate a second time than individuals initially stimulated to donate by the physical attractiveness of charity solicitors.

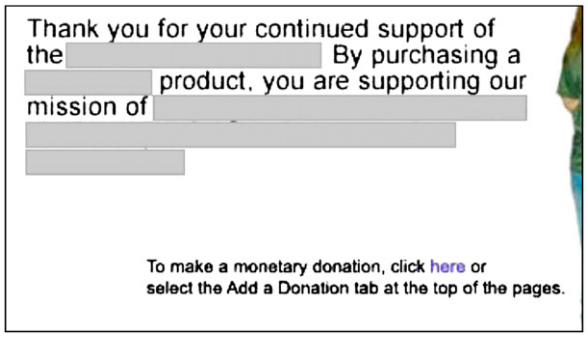

Fig. 1. Store central graphic in Null treatment.

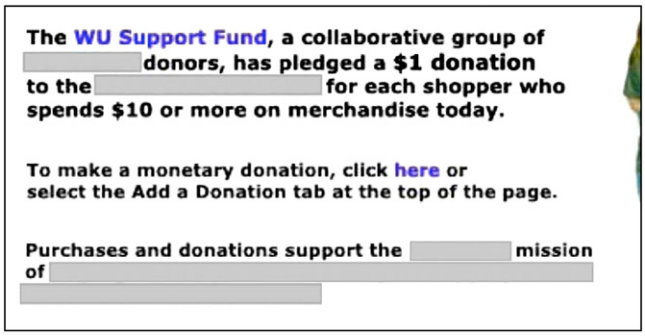

Fig. 2. Store central graphic in P:1 treatment.

advertising banners. Around the time of the experiment the store received approximately 60,000 visits per month yielding 1200 orders. A typical store customer purchases two or three items during a visit, presumably for household use, but some customers represent firms or organizations that require a large quantity of items. During spring 2007 excluding the weeks of the experiment, the median order size was $28 and approximately 97% of orders included at least $10 of merchandise. During the same period, a limited number of consumers (about 1%) included donations with their orders.

Consumers who enter the store are shown a large graphic which displays a promotional message. The store's front page always includes a reminder that consumers' orders support the NPO's charitable mission. To the side of the front page's central graphic there is a list of product category links that take consumers to summary pages which display the pictures and prices of items within the product categories. From these summary pages, a consumer can click on an individual item to see a detailed picture, product description, and price, and the consumer can elect to place one or more units of the item in his shopping cart.

2.2. Experimental treatments

We conducted the experiment from March 6, 2007 until April 15, 2007.[4] During this period, each store visitor had a cookie placed in his or her internet browser to assign an experimental treatment and identification code. We use this code to track store visitors who return multiple times within the experimental period.[5] As long as the visitor's cookie was not deleted from his browser, each time he returned to the store he observed the same experimental treatment. Consumers who deleted cookies from their browsers or visited the store from multiple computers (e.g. home and work) were assigned a new cookie and likely placed into a different experimental treatment. Unfortunately, we cannot measure the number of consumers who observed multiple experimental treatments due to this characteristic of cookies. In the discussion below we describe aspects of the study and data that indicate this is unlikely to be a major problem.

There are two components to each experimental treatment. First, each consumer received one of six messages regarding the charitable value of purchasing at the NPO's online store. Second, each consumer was assigned either the regular online store prices or a set of discounted prices. In total, there were 12 treatments in the experiment (6 messages × 2 price levels), and each was assigned with equal probability. During the experimental period the NPO suspended all other price promotions for its products.

2.2.1. Donation messages

Consumers' messages regarding charitable value were delivered through the central graphic of the store's front page. In the control case ("Null"), the central graphic displayed text indicating that store activity benefits the NPO's charitable mission. In Fig. 1 we display the relevant portion of this central graphic, with identifying information about the NPO obscured. The extended section of obscured text is a description of the NPO's mission, and the partial image on the right of the Figure is a young girl wearing NPO apparel. The first two non-Null messages promised consumers that an additional donation would be made if they purchased at least $10 in merchandise at the online store. The donation values were $1 and $5, which we identify here as "P:1" and "P:5," respectively, and as "P:δ" when referenced as a pair. The donations were pledged by the semi-anonymous "WU Support Fund" (WUSF) and were paid by the researchers. See Fig. 2 for the relevant portion of the P:1 graphic. The P:5 graphic is identical except for the donation amount. We include the word "merchandise" in the message so consumers were aware that shipping charges would not be included in the comparison of their order value to the $10 threshold. In the final set of messages, consumers were told that if they purchased $10 or more in merchandise plus donated $10 or more at the store, the WUSF would donate $1, $5, or $10. We identify these treatments as "$P+D$:1," "$P+D$:5," and "$P+D$:10," and as "$P+D$:δ" when referenced as a group. In Fig. 3 we display the central graphic for the $P+D$:1 message. The $P+D$:5 and $P+D$:10 graphics are different only in their donation amounts. Each non-Null treatment included a block of fine print that described the WUSF ("a group of private individuals committed to assisting the [NPO] in achieving its mission"), stated that the donation offer was limited to one per customer for the duration of the promotion, and specified that there was no maximum total donation from the WUSF.

In addition to the WUSF donation pledges on the front page's central graphic, we placed condensed versions of each message near the top of the store's product-level pages. In Figs. 4 and 5 we display these condensed messages for the P:1 and $P+D$:1 treatments. Consumers who were assigned the Null treatment received no message in this space. We included these reminders in case a consumer clicked past the store's front page without reading the text in the central graphic.

In order to handle potential consumer confusion about the donation messages, the NPO provided a phone number for customer service. The NPO's customer service employees were prepared to give callers additional information about the donation pledges, but the NPO reported that no consumers called to request this information.[6] The absence of phone calls provides some evidence that consumer exposure to multiple experimental treatments is not a serious concern.

[4] The experiment's starting date and duration were decided jointly with the NPO, and was limited in part by the NPO's desire to vary the store's promotional offers, which generally were suspended during the experiment.

[5] See Chen et al. (2006) on the variety of methods available for online field experiments.

[6] Customer service representatives were also instructed offer a consumer the discounted set of prices if the consumer saw both price levels and was confused by the difference. No consumers called with questions about prices.

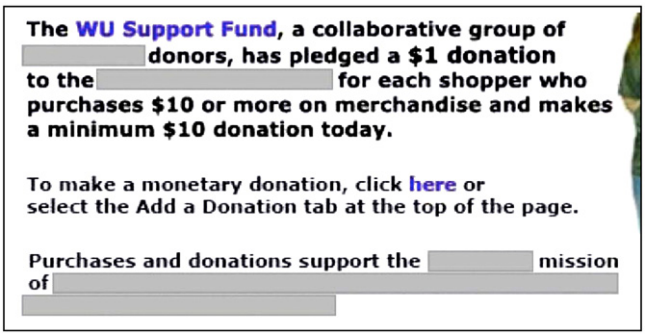

Fig. 3. Store central graphic in P+D:1 treatment.

The $10 threshold for consumers' purchases is central to the discussion below. In designing the experiment, we anticipated focusing on consumers' extensive margin responses. The presence of some low-priced items ($5–$10) at the store led us to create the threshold so that WUSF donations of $5 in *P*:5 did not appear overly large relative to the prices of some items at the store. For consumers who selected one small item only, a large WUSF donation could reduce comparability between our study and more general firm practices of bundling private and public goods. In addition, if the pledged donation appeared suspiciously generous under certain circumstances, this could reduce the credibility of the WUSF messages for *all* consumers, regardless of their interest in small items. While we did not set the $10 threshold optimally to generate information on how consumers interpreted WUSF messages, we demonstrate below that our data are sufficient for this purpose.

To understand the benefits of using the lump-sum WUSF donations rather than another mechanism that generates charity value, consider a few alternatives. First, it is possible to simply tell consumers that a $30 item includes an implicit donation of either $3 or $6 to the NPO's mission. One potential problem is that only one (at most) of the two donation amounts can be true without outside funding. An additional problem is that without describing the presence of outside funding, consumers may make different inferences about the underlying quality of the product being sold — in this example that the item's cost is either $27 or $24. Second, we might have used varying percentage donations from the WUSF, for example that the WUSF would donate 1%, 5%, or 10% of a customer's order value. While this scheme has the benefit of resembling some firms' actual pledges of donations in proportion to sales, paying pledged WUSF donations in these cases could be prohibitively costly due to a small number of store customers who place very large orders (several thousand dollars). In Section 3, once we have introduced a model of consumers' choices, we describe additional benefits of the lump-sum WUSF donations in allowing us to distinguish among possible explanations for consumers' responses to the experiment.

2.2.2. Price variation

To assess how price sensitivity changes with the strength of a product's charity association, we randomly assigned consumers to either the regular (non-sale) store prices or discounted prices. Of the 45 items offered at the store, we received permission to adjust prices on 20 items, which account for about half of the store's sales. When the discount was applied, the prices of these 20 items were reduced by 8–20% relative to the control. Focusing on items that were potentially-discounted during the experimental period, the (weighted) average regular price was $32.20 and the average discounted price was $28.54. Variation in discounting percentages was due to suggestions from the NPO that we change prices in whole-dollar increments while also selecting price values that would not appear out-of-place on the store website. Products with initial prices above $30 had relatively small percentage price discounts, while those with initial prices below $30 had discounts that averaged 16%. The 25 items that were not subject to discounting had an average price of $15. Other than the price levels, the discounted prices had the same appearance as the regular prices. There were no additional messages about special sale prices.

Fig. 4. Reminder graphic for P:1 treatment.

Fig. 5. Reminder graphic for P+D:1 treatment.

3. Incentives generated by the experiment

We now provide a discussion of how consumers might respond to the experimental treatments. We first consider what would occur if consumers' preferences and information about the NPO are stable and consumers place no value on potential WUSF transfers that are not made. We do not consider all permutations of the treatments here, but instead provide an overview of some important patterns that would appear in the data. We then comment on how consumers' actions could be different if a treatment's main impact is to generate information about the value of supporting the NPO. In the contrast to the more conventional incentive model, we describe the information's impact as changing consumers' preferences over revenue transferred to the NPO.

Assume that there exists a consumer population that receives utility from: units of a product (x) offered by the NPO, transfers of their own (d) to the NPO, transfers made by the WUSF (δ_A) to the NPO, and a numeraire (z). We write δ_A to represent actual transfers to the NPO as opposed to the offered transfer (δ) in a WUSF message. When a consumer meets the terms of a WUSF message, $\delta_A = \delta$ but otherwise $\delta_A = 0$. All consumers share the symmetric income level y and pay a price p per unit of x. To capture the main features of the merchandise component of the WUSF orders, we assume that $x \in \{0, 1, 2\}$ with an order of $x = 1$ being large enough to trigger a WUSF donation. The variables d and z may take any non-negative value so long as they, together with x, satisfy the consumer's budget constraint, $px + z + d \leq y$.

Let $v_i(x) = \theta_i v(x)$ represent consumer i's utility from receiving x units of the NPO's product. Consumers vary only in their value of θ_i, which we assume is distributed according to F on the interval $[0, \infty)$ with positive density throughout the support. The function v is increasing, concave, and has $v(0) = 0$. Consumers' utility from transfers to the NPO is captured by the function $w((p-c)x, \delta_A, d)$, where c is marginal cost and $(p-c)x$ is "profit" to benefit the NPO's charitable mission. Assume that w is increasing and concave in each argument, and $w(0, 0, 0) = 0$. Let w_j represent the derivative of w with respect to its jth argument, and let w_{jk} be the cross-partial derivative between arguments j and k. If a consumer views NPO revenue from different sources as substitutes, then $w_{jk} < 0$ across pairs of distinct arguments. We combine the benefits in v, w, and z linearly to form consumer i's utility function: $U_i = \theta_i v(x) + w((p-c)x, \delta_A, d) + z$. We assume that y is large enough to permit an interior solution in x and d, if desired by the consumer, and that v, w, and F lead to a positive share of consumers selecting each feasible x value. Finally, to further

simplify consumers' choice problems under the P:δ messages, assume that $w_3(.,.,0)<1$ so that consumers always set $d=0$ when not prompted by a $P+D$:δ message to make their own donation. This assumption is realistic for likely store customers under the empirical donation patterns within the store and across NPO donation platforms (see Section 4).

The simple form of preferences implies that each treatment will generate two threshold values of θ that divide the population by their choices of x. For a consumer population that receives message m and faces undiscounted prices, we write θ_j^m to indicate the value of θ that separates consumers who buy j units from those who select $j-1$ units. Under the Null treatment (with $\delta_A=\delta=0$), these values are θ_1^N and θ_2^N. Now consider how consumers around θ_1^N and θ_2^N with fixed preferences in v and w respond to WUSF messages. When consumers are assigned a P:δ message, the additional utility from $x\geq 1$ through w will draw additional consumers into the market at the extensive margin. This effect is stronger with a P:5 message than P:1, thus we expect the share of store visitors who purchase to be greater with P:1 than under the Null, and larger still with P:5. The P:δ messages, however, may reduce demand at the intensive margin. For a consumer with $\theta=\theta_2^N$, increasing δ_A from zero could reduce the marginal benefit of his own implicit donation through x if $w_{12}<0$. If instead $w_{12}=0$ because consumers do not view the WUSF's donations and their own as substitutes, then consumers who would have purchased under the Null should not change their choices of x.

The $P+D$:δ treatments are designed to stimulate consumers' donation decisions while they are in the NPO store. The pledged WUSF donation effectively lowers a consumer's "price" ($10) of a $11, $15, or $20 donation. This price reduction for d may expand demand for x at the extensive margin for consumers near θ_1^N who find the prospect of $x>0$ and $d>0$ appealing at the new price of d. Just as with the P:δ treatments, a consumer with $\theta=\theta_2^N$ who chooses $d>0$ to trigger $\delta_A>0$ may strictly prefer $x=1$ to $x=2$ because the consumer's utility benefit of generating donations through x falls if $w_{13}<0$. The impact of δ_A can increase the likelihood of a reduction in x if $w_{12}<0$ as well. Finally, we note that consumers who choose $d=0$ under a $P+D$:δ message should have the same distribution of x values as they would under the Null.[7]

The implications are less clear for interactions between price changes and WUSF offers. For consumers at the extensive margin, a reduction in p leads to a reduction in the threshold value of θ_1^N that is proportional to $-(1-w_1)$. This term includes the direct benefit of a lower price for x and an offsetting effect due to the reduction in the consumer's implicit donation. WUSF transfers of δ_A could reduce the dampening effect of w_1 (thereby increasing price sensitivity) if $w_{12}<0$, and additionally so if $w_{13}<0$ when a consumer contributes $d>0$. Price sensitivity across treatments is also affected by potential differences in the densities of marginal consumers around the different messages' values of the threshold θ_1^m. Similarly, price sensitivity on the intensive margin is uncertain without further assumptions on w and F. In considering the difference in w for $x=1$ versus $x=2$, a consumer may find that this difference increases or decreases with a reduction in p, depending on w's curvature.[8] Changes to w_1 through non-zero cross effects involving δ_A and d further complicates this comparison. Given the variety of possible price effects across treatments, we leave the determination of price sensitivity as an empirical question.

There are several ways the model could be extended to account for consumers' repeated choices over time. One possibility is that U is defined for a period longer than the experiment. Under this approach, if the WUSF messages induce consumers to select large values of x or d during the experimental period, then their marginal benefit from implicit or direct donations later may be relatively low. Alternatively, we could interpret U as a flow utility for each round of x and d choices, and consumers maximize discounted lifetime utility subject to an intertemporal budget constraint. The WUSF messages would introduce a shock to the consumer's incentives over x and d during the experimental period, but in later periods the consumers' choices could revert to their levels under the Null, or perhaps decline if the consumers have self-imposed limits on their total charity expenditures, for example. In either extension of the model to choices over time, we expect that consumers with WUSF messages and stable preferences would have post-experiment store purchases that are no greater than those of consumers with the Null message.

In contrast to the responses as described above, where consumers with stable preferences respond conventionally to the financial terms of the experimental treatments, it is possible that the NPO store visitors process the WUSF messages differently. As in Shang and Croson (2009), consumers who receive information about other agents' actions may update their own preferences regarding a charitable organization or its products. It is possible that no treatments, some treatments, or all have this effect. This updating may occur within a broader model in which consumers believe they hold incomplete information about the NPO's quality, and new information prompts an updating of their quality inferences. In a "reduced form" of this quality inference model, the WUSF message may change the properties of w, especially through the benefit consumers receive from their own implicit (x) or direct (d) donations, regardless of whether they generate the donation δ_A.[9]

We use the experimental data to determine the dominant interpretation (stable or shifted preferences) of each WUSF message. Two aspects of the experiment and model allow us to make this evaluation. First, a consumer whose value of w_1 has increased with exposure to a P:δ message is both more likely to purchase and more likely to select $x=2$. This follows from a simple reduction of θ_1^m and θ_2^m thresholds when implicit donations have greater value. By contrast, the stable preferences model suggests that the share of consumers selecting $x=2$ remains unchanged or falls. This opportunity to distinguish among theories using intensive margin responses would have been impossible had we created the P:δ messages with a percentage donation (e.g. 5% of a consumer's total order); an increase in x could be ascribed to a shift in the marginal financial incentives or a change to w_1.[10] Second, if consumers predominantly respond to $P+D$:δ messages by updating their preferences in w, then they may increase their choice of x regardless of whether they satisfy the terms for generating $\delta_A>0$. Consumers with stable preferences who receive $P+D$:δ messages and choose $d=0$ should not change their value of x along any margin, while those who donate and trigger $\delta_A>0$ should be no more likely than consumers with the Null to choose $x=2$.

Similar distinctions may be noted in consumers' responses to price variation and their activity beyond the experiment. If the WUSF message shifts a consumer's preferences in w, the increased value of implicit donations through x could reduce price sensitivity rather than increasing it or keeping it fixed. Likewise, the stable preferences interpretation of the model predicts that the post-experiment expenditures of consumers with WUSF messages should fall, either to the Null level or below it, while an increase in w leads to greater later activity. Such an increase in long-term activity may be perceived as habit formation, but here this effect would have a specific

[7] If NPO revenue is perceived as a public good, the consumer may reduce his choice of x if the marginal value of his implicit contribution is reduced by the belief that other consumers will change their donation and purchasing behavior (in addition to revenue from the WUSF).

[8] For example, if $w=\log[(p-c)x]+g(\delta_A, d)$ for some appropriate g, then the difference in w values for $x=1$ and $x=2$ is unaffected by changing p, while other w specifications could cause the difference to increase or decrease with p.

[9] Alternatively, it is possible that a WUSF message prompts consumers to update their quality inferences on x itself, which would be reflected in v. We do not favor this interpretation because our WUSF messages suggest that an outside group is interested in supplementing the NPO's revenue, which appears in w for consumers.

[10] This is a challenge in interpreting the results in Karlan and List (2007), who employ a proportional donation-matching scheme.

interpretation tied to the information in WUSF messages and its effect on consumers' tastes.

4. Data

4.1. Reports from the online store

After the experiment ended, the NPO provided data on activity at the online store. These data capture activity at two levels. First, we observe basic information about every visit ("session") to the online store. For each session, we observe the time and date it was initiated, the visitor's Internet Protocol (IP) address, and the visitor's assignment into an experimental treatment.[11] Through the cookie stored in a consumer's browser, we are able to track consumers' repeat visits to the store. The second level of data is for consumers' purchases. For each consumer who placed an order at the store, we observe all session-level details listed above, the items selected and prices paid, any donation activity within the store, and the consumer's billing zip code. We use consumers' email addresses (which are entered during check-out) as an additional way to link individual consumers' activity across multiple transactions at the store.

The personal information, browser cookies, and identification codes of online store visitors were not used by the NPO or the researchers for any purpose other than the academic study we describe in this paper.

4.2. Consumer activity at the store

During the experimental period of March 6 to April 15, 2007 there were about 6200 sessions initiated for each of the 12 treatments. These initiated sessions represent approximately 5800 consumers per treatment that we examine in this paper. Some of the difference between the total sessions and studied consumers is due to consumers making multiple visits to the store before purchasing.[12] On average, consumers who did not order visited the site 1.04 times during the experimental period, while consumers who ordered visited 1.61 times. Other store visits are eliminated because of short-term disruptions in our treatment-assignment mechanism. In all, the sessions we study led to orders placed by 1392 unique consumers. For each consumer who purchased multiple times within the experimental period (and under a single treatment), we aggregate all orders into a single observation. This affects only 36 consumers, and our empirical results are robust to whether these consumers' orders are disaggregated or dropped altogether.

About 2% of store visitors placed orders, with an average order size of $86.91. This implies that each unique store visitor generated $1.74 in revenue for the NPO. These values and similar statistics in this paper include only merchandise and exclude shipping charges. Due to a small number of very large orders, the median order is $37, considerably below the mean. In general, the largest orders are placed on behalf of firms rather than households, but no perfect method exists to identify which orders are from firms. To focus on households' choices, which are more relevant for our objectives and the literature to which we are contributing, within each WUSF message (including the Null) we drop orders above the 95th percentile of order values.[13] After eliminating the top 5% of non-zero order values, the average order is $50.38 for a per-visitor average of $0.96. Eliminating very large orders also leads to a substantial reduction in the variance of consumers' order sizes. Across all treatments, the coefficient of variation (σ/μ) falls from 3.7 to 0.9 when we drop the largest 5% of orders.

In the top panel of Table 1 we display the number of visitors and their order values for the six experimental messages, including the Null. Regular and discounted price treatments are combined within messages in this portion of the Table. Consumers' purchase probabilities were marginally greater in most of the non-Null treatments, but none of the differences were large. The more striking difference is between the average order values in Null and non-Null messages. Revenue per store visitor was $0.63 greater when consumers received a non-Null message. This is due to a difference of $30.56 per customer conditional on an order. The difference in order sizes between the Null and non-Null messages is $8.90 when we drop the largest 5% of orders from each subsample. With and without dropping large orders, the median order value is $40 for orders placed under each WUSF message except P:5, while both the P:5 and Null messages have median values just under $30.

The average order values across messages generally follow the pattern in the median order values. Excluding the largest 5% of orders, the order value per visitor for the P:1 treatment exceeds the Null by 30%, while the same statistic for P:5 is 5% greater than the Null. Each $P+D$:δ message has an average value per visit that is substantially larger than the Null, varying from 13% to 32% greater. In Fig. 6 we display the cumulative distribution functions for three groups of orders. The order distribution under the Null (blue solid line) is strictly to the left of the orders placed under a P:1 or $P+D$:δ message (green dashed line). This pattern is repeated if P:1 or any $P+D$:δ message is compared individually to the Null. The distribution of P:5 orders (red dotted line) follows the other non-Null treatments until about $25, beyond which it follows the distribution of Null message orders for greater values. We add the dashed line at $10 to ease examination of the distributions relative to the threshold for triggering a WUSF donation.

Despite increases in their order values, consumers in the $P+D$:δ treatments predominantly chose to omit donations from their orders. This regularity led to a smaller total donation ($1636) from the WUSF to the NPO than we expected, but the NPO benefited strongly in other ways from the $P+D$:δ treatments. Each $P+D$:δ message generated merchandise revenue that was, on average, about $2600 greater than the Null's revenue over the experimental period. As shown in Table 1, about 4% of consumers who ordered under a $P+D$:δ treatment made a donation, with the greatest number donating within the $P+D$:10 message. Consumers with $P+D$:δ messages who donated had orders that average $4.65 less than those of consumers who did not donate. It is important to emphasize that when consumers in $P+D$:δ treatments did not make their own donation, then no WUSF donation was triggered. This regularity, along with the stronger consumer reaction to P:1 than P:5, presents a significant puzzle within the data. As we describe in Section 3, this response is difficult to explain with a single model of consumer behavior but more reasonable if some WUSF messages effectively shifted w_1.

[11] While an IP address may reveal a consumer's location, our IP data are insufficient for adding information on consumers' geographic locations to our empirical analysis. The IP address provides the location of the consumer's Internet Service Provider, which can be in a different city or state that the consumer himself. In unreported analysis, we compared zip code-level demographics for consumers who purchased under our various treatments, and we found that consumers' locations were not significantly related to their choices under the WUSF messages.

[12] We investigated sessions (and subsequent orders) that were initiated in rapid succession from the same IP address, with the store visitors apparently deleting their browser cookies between each session. This may be due to testing of the website by the NPO or ourselves, or it may be consumers who realize that the store appearance changes each time they delete cookies and begin a new session at the store. (Site testing never leads to orders at the store.) About 4% of store orders follow from a session associated with this sort of activity, and these store orders are distributed fairly evenly across experimental treatments (including price discounting). This suggests that few store customers were behaving opportunistically with regard to inserting themselves into experimental treatments with more favorable terms. None of the paper's results change if we drop observations by consumers who appeared to delete cookies between store visits.

[13] The 95th percentile is $165 for consumers with the Null, and for individual WUSF messages the cut-off averages $250. By defining the threshold separately for each group of data, we minimize bias due to differences across experimental treatments in what share of observations within the treatment would be above a single threshold.

Table 1
Summary statistics.

	N consumers		Purchase probability	Avg. revenue per visitor		Avg. revenue per customer		Median order	Number of donations
	Visits	Orders		All	Drop large	All	Drop large		
Panel 1: Activity by message									
Message type									
Null	11,386	226	0.0198	1.217	0.811	61.31	42.92	28.00	2
Any non-Null	58,075	1166	0.0201	1.845	0.991	91.87	51.82	39.92	35
P:1	11,893	232	0.0195	1.700	1.057	87.14	56.81	39.98	2
P:5	11,734	243	0.0207	1.649	0.854	79.63	43.32	29.90	3
P + *D*:1	11,464	232	0.0202	2.528	1.074	124.90	55.64	40.00	7
P + *D*:5	11,487	220	0.0192	1.434	0.913	74.86	50.11	40.00	7
P + *D*:10	11,497	239	0.0208	1.925	1.061	92.58	53.46	39.95	16
Panel 2: Activity by pricing treatment									
Discount?									
No	35,098	686	0.0195	1.635	0.881	83.66	47.35	37.90	17
Yes	34,363	706	0.0205	1.851	1.012	90.09	51.79	37.00	20
Panel 3: Supplemental data									
Not during experiment	31,342	721	0.0230	1.470	0.921	63.91	42.16	28.00	9

Notes: "Drop large" statistics are calculated after eliminating orders above the 95th percentile of non-zero orders within the subsample utilized in each row. The supplemental data are from visits to the store in March and April that occurred outside of the experimental period.

In Panel 2 of Table 1 we display average order values within the price treatments. As expected, lower prices led to both an increased purchase probability and an increased order size per store visitor. In total, store visitors who received discounted prices spent about $4000 more than the $30,900 in revenue generated by visitors who were assigned the normal store prices. These revenue figures and those in the Table understate the magnitude of consumer response to the price reductions. Because of discounting, the increase in quantity exceeded the increase in revenue; we account for this in the analysis below. An additional factor that complicates interpretation of Panel 2 is that the statistics include products that were never discounted. When we focus on the products that were included in the discounting portion of the experiment, we find that the chance of a store visitor purchasing at least one of these products increases from 1.06% without a discount to 1.26% with the discount. This is a greater proportional change in purchase probability than when all merchandise orders are included.

Our experimental structure included equal division of consumers into the Null and each non-Null message. For some of the analysis below it is useful to compare a larger group of Null-like consumers to those who received a non-Null message. To do this, we augment the experimental data with store data from March and April outside of the experimental period.[14] Consumers who purchased from the store during these portions of March and April saw store graphics and prices similar to those in the Null portion of the experiment. Store activity during the supplemental period was similar to activity under the Null. See Panel 3 of Table 1 for summary statistics on the supplemental period. Of 31,342 visitors during this period, 2.3% of consumers placed an order (2% for the Null), and the average order value in the lower 95% of orders is $42.16 ($42.92 for the Null). Differences in the percentage of visitors who purchase may be due to promotions that ran during the supplemental period or general consumer shopping patterns during these portions of March and April.

4.3. Actions after the experiment

We supplement the data on consumers' choices in March and April 2007 with information on the same consumers' interaction with the NPO between May 2007 and March 2008. Due to the anonymity of consumers who visit the store and leave without purchasing, we are limited to consumers who purchased during the experiment and supplemental periods.[15] We study two aspects of these consumers' choices. First, we collect data on the consumers' transactions at the online store. Second, we received information on donation dates and amounts from the NPO database of online donors, which does not include donations made at the store. Of the 1392 consumers observed during the experiment, we identified 86 consumers who were active with the NPO outside of the experimental period. Most of these consumers (69) were repeat customers of the store but not donors, and only 1 consumer appeared in both the follow-up store data and the donation data.

In examining a consumer's actions after the experimental and supplemental periods, we focus on transactions that occurred between 60 and 300 days after the consumer's last action during March and April 2007. We do this so that there is no overlap between subjects in the experimental and supplemental periods.[16]

In Table 2 we provide summary statistics on consumers' purchases outside of the experimental and supplemental periods. Consumers who ordered multiple times after the experiment have their orders summed. In the Table and all related analysis we omit observations that are in the top 5% of orders during the experiment or supplemental period. Consumers who saw the Null message or were in the supplemental group returned at a rate of 2.6%, while 4.0% of consumers who received a WUSF message returned to purchase. Conditional on returning to the store, consumers who received WUSF messages placed substantially larger orders.

Considerably fewer store consumers appear in the NPO's roll of online donors. This donor database contains over 2 million names, and it accounts for a large share of the organization's donation revenue. Despite this, only 24 of the 2113 customers from March and April 2007 appear in the donor database, with 18 consumers making donations after April 2007. There were 12 consumers (1% of 1169) who saw a WUSF message and donated later, and 6 consumers (0.7%

[14] We exclude consumers who visit the store on April 16–17 to minimize the impact of consumers who first visited the store during the experimental period and were exposed to one of our WUSF messages. Due to data limitations, we also exclude consumers who entered the store on March 1 and after April 27.

[15] After the experimental period no new cookies were placed in consumers' web browsers, and the cookies that were placed during the experiment expired soon after its conclusion.

[16] For example, we do not want a consumer who purchases on March 10 and returns to the store on April 20 to be part of both the experimental and supplemental groups. But if the March 10 consumer's actions are omitted from the analysis until 60 days after purchase, then a consumer who first arrives at the store on April 23 should be treated in the same way.

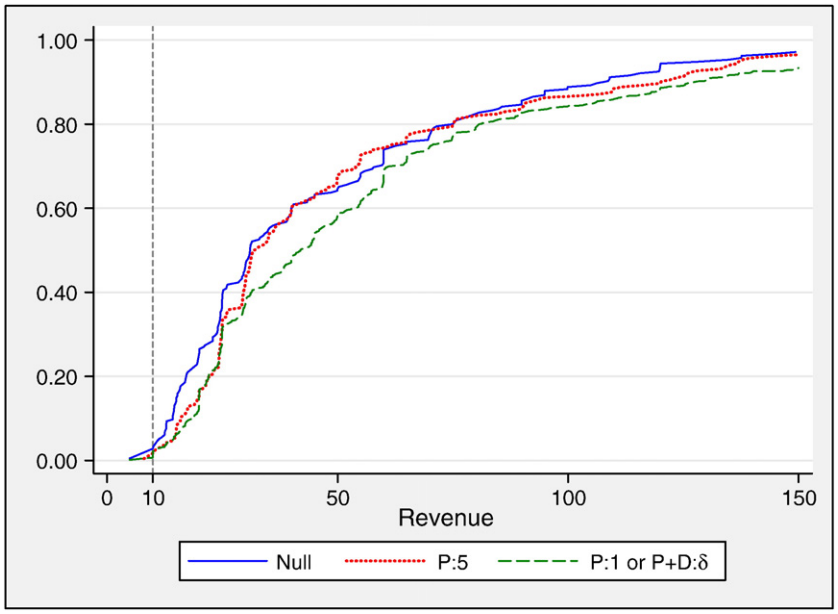

Fig. 6. Cumulative distributions of order revenue.

of 947) who donated after seeing the Null message or are part of the supplemental group. While we forgo further empirical analysis of donation patterns because of the small number of observed consumers, the scarcity of these consumers may be considered a result in itself. This suggests that the NPO's collections of store customers and donors are largely distinct, and charity-oriented promotions at the online store are likely to have little impact on donation activity. The WUSF messages' sizable impact on order sizes had little opportunity to affect donation revenue.

5. Detailed analysis and testing

We now examine more closely the impact of the experiment on behavior. In evaluating the overall impact of the experiment, we use session- and order-level observations to differentiate between the behavioral theories in Section 3 that could describe consumers' responses to the message treatments. We also examine consumers' price sensitivity and whether exposure to WUSF messages affected behavior after the experiment.

Our primary dependent variable is r, the online store revenue generated by an individual consumer. We consider both r at the level of all store visits, which implies $r=0$ for many consumers, and r conditional on a merchandise purchase ($r>0$). While the NPO's charitable mission benefits from store "profit" rather total revenue, consumers' demand responses to WUSF messages are demonstrated effectively with r. Order revenue allows us to account for the wide variety of objects for sale at the store, so that Section 3's variable x is more similar to r than x is to the quantity of items purchased. We do not observe the store's cost of filling orders.

When evaluating the impact of experimental treatments on consumers' choices, the main econometric model is

$$r_i = \alpha + \beta T_i + \varepsilon_i, \quad (1)$$

where T_i is a vector of dummy variables that identify the treatment or group of treatments (e.g. any WUSF message) to which consumer i was randomly assigned. The error term ε_i captures both variation in consumers' tastes for the NPO's products and in their responses to non-Null treatments. To minimize distributional assumptions on ε_i and ease interpretation of the estimated coefficients, we generally estimate the parameters of Eq. (1) with Ordinary Least Squares (OLS) while computing robust standard errors to account for heteroskedasticity. In some analysis we adapt Eq. (1) to account for log transformations of r or binary dependent variables that describe consumer choices.

Table 2
Post-experiment purchases.

	Consumers with later orders		Revenue from later orders	
Consumer group	N	Share	Mean	Median
Null	4	0.0185	37.08	39.25
P:1	9	0.0411	111.78	74.85
P:5	5	0.0215	154.10	95.72
$P+D$:1	11	0.0507	108.48	43.95
$P+D$:5	13	0.0607	132.89	59.95
$P+D$:10	6	0.0265	38.58	32.00
Supplemental data	19	0.0278	52.35	32.95
Null + supplement	23	0.0256	49.69	32.95

Notes: Consumers included in this Table are in the lower 95% of orders during the experiment and supplemental periods. All transactions occurred between 60 and 300 days after the consumer's last action at the store during March and April 2007.

5.1. Impact of the experiment on order revenue

We begin by presenting a set of session-level results in Specifications 1–3 of Table 3. For Specification 1 we regress r on a simple indicator of whether a store visitor saw a WUSF donation pledge. The model's constant recovers the average transaction value under the Null, 0.81, as we reported in the summary statistics of Table 1. Consumers who received any non-Null treatment spent an average of $0.18 more than consumers with the Null. In Specification 2 we divide the WUSF messages by their individual content, and we find that the two messages with $5 WUSF donations had the smallest impacts on store revenue; their coefficient estimates are positive but statistically insignificant. The estimates for the remaining messages are around $0.25, and each is significantly different from zero. In Specification 3 we show that discounted prices have a marginally significant impact

Table 3
Impact of treatment messages on revenue by visitor and conditional on order.

Dependent variable	Revenue per visitor			Order revenue		log(Order revenue)	
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.811***	0.811***	0.880***	42.92***	42.92***	3.427***	3.427***
	(0.0719)	(0.0719)	(0.0442)	(2.468)	(2.472)	(0.0562)	(0.0563)
Any message	0.180**			8.899***		0.197***	
	(0.0823)			(2.847)		(0.0610)	
P:1		0.245**			13.89***		0.247***
		(0.122)			(4.500)		(0.0796)
P:5		0.0425			0.403		0.0663
		(0.102)			(3.378)		(0.0740)
$P+D$:1		0.262**			12.47***		0.216***
		(0.124)			(4.412)		(0.0807)
$P+D$:5		0.101			7.189**		0.228***
		(0.107)			(3.636)		(0.0756)
$P+D$:10		0.250**			10.54***		0.235***
		(0.118)			(4.019)		0.216***
Discounted pricing			0.132*				
			(0.0686)				
N	69,394	69,394	69,394	1325	1325	1325	1325

Notes: Omitted category is Null message or No Discount, as appropriate. Each model drops the top 5% of consumers within treatment message or, for Specification (3), within price treatment. Robust standard errors are in parentheses.
***Indicates 1% significance; **indicates 5% significance; and *indicates 10% significance.

on per-visitor store revenue. Below we analyze some interactions between WUSF messages and the price discounting treatment.

The results in Specifications 1–3 are driven by increases in spending conditional on purchasing from the store. Additional unreported analysis shows no significant impact of WUSF messages on the probability that a consumer placed an order, regardless of the level of aggregation across WUSF messages. This may have occurred because the positive impact of WUSF pledges on some visitors was balanced by other visitors responding negatively to the messages, perhaps because consumers found the messages confusing or unappealing due to their text-heavy aesthetic. Alternatively, the bulk of store visitors may have been interested in browsing only, making them very unlikely to move across the extensive margin to purchase. Among consumers who ordered, the central remaining question is why they had a stronger positive response to the (untriggered) $P+D$:δ messages and *P*:1 message than the relatively rich *P*:5 message.[17] We suspect that the *P*:1 message and all three $P+D$:δ messages acted only to shift consumers' perceptions of the store as an effective way to support the NPO, as if w_1 increased in Section 3's model. The specific details of these WUSF offers were either unnoticed or ignored because of low stakes (*P*:1) or general disinterest on the part of NPO store customers in making cash donations ($P+D$:δ). The *P*:5 message, by contrast, had higher stakes and required only a decision over purchases — a choice that consumers may have been better prepared to consider when they entered the store. We conjecture that these properties of the *P*:5 message drew consumers' attention to the specific details of the pledge, which is closer to the model in Section 3 with unchanging preferences for NPO revenue.

To explore these ideas further, we now turn to examining order revenue conditional on an order. This approach allows a greater focus on order sizes and requires minimal sacrifice in ignoring the extensive margin. In addition, the stable versus changing preference interpretations offered in Section 3 have their sharpest differences in order sizes conditional on a positive order. In Specifications 4 and 5 of Table 3 we display the impact of each message on order revenue. Within specific WUSF messages, the *P*:1, $P+D$:1, and $P+D$:10 treatments have significant impacts on order size, and the magnitude of these effects are similar to each other (between $10.54 and $13.89). The impact of the $P+D$:5 message is positive ($7.19) but smaller. The *P*:5 message stands out with an estimated coefficient that is nearly zero ($0.40). In Specifications 6 and 7 we repeat the analysis but use $\log(r)$ as the dependent variable, which reduces the impact of very large orders on the orders' variance. Each WUSF message other than *P*:5 has a positive and significant impact on $\log(r)$, with the coefficient magnitudes indicating an increase in order revenue of about 25%. The impact of the *P*:5 message, by contrast, is just under 7% and insignificantly different from zero.

With Table 3 we established that the mean responses to the *P*:5 message are different from the responses to other WUSF messages, which in turn are similar to each other. Following either of the theories from Section 3, we would expect consumers to have responded differently to *P*:5 than the Null, but the results in Table 3 indicate no difference. As we suggest with Fig. 6 and the model in Section 3, differences across treatments are characterized by differences in the whole distribution of orders rather than the mean. We examine this directly with a series of

Table 4
Message impact throughout the distribution of orders.

	Quantile				
	0.10	0.20	0.30	0.50	0.75
	(1)	(2)	(3)	(4)	(5)
Constant	10.00***	15.00***	20.00***	28.00***	64.95***
	(0.677)	(1.617)	(0.797)	(3.771)	(6.660)
P:1	5.95***	5.00***	2.94	12.00**	14.95
	(1.499)	(1.618)	(2.025)	(5.574)	(13.30)
P:5	4.85***	4.95**	0.00	1.90	−4.95
	(1.436)	(2.034)	(2.114)	(4.898)	(9.787)
$P+D$:1	2.95*	5.00**	4.95**	12.00**	10.05
	(1.779)	(2.003)	(2.150)	(5.073)	(9.962)
$P+D$:5	6.95***	5.00**	5.00**	12.00**	0.00
	(1.463)	(2.303)	(2.022)	(4.837)	(10.10)
$P+D$:10	5.95***	5.00***	2.00	11.95**	7.05
	(1.037)	(1.646)	(1.967)	(5.220)	(11.25)
Alternative specification (see notes)					
Any message but *P*:5	5.95***	5.00***	4.95***	12.00***	8.05
	(0.936)	(1.460)	(1.316)	(3.773)	(7.860)
N	1392	1392	1392	1392	1392

Notes: Standard errors, calculated by bootstrapping with 500 repetitions, are in parentheses. The alternative specification also includes a constant and an indicator for *P*:5. Point estimates for the constant and the *P*:5 coefficient are identical to those in the table, and differences in standard errors are minimal.
***Indicates 1% significance; **indicates 5% significance; and *indicates 10% significance.

[17] We show below that the $P+D$:5 message is more naturally grouped with the other $P+D$:δ messages than with *P*:5.

Table 5
Further analysis on order characteristics.

Dependent variable	Prob(Order ≤ 10)		log(Avg. price)	Item quantity	Prob(Item quantity ≥ 2)	
Includes sub-$10 items?	–	–	–	–	Yes	No
Specification	(1)	(2)	(3)	(4)	(5)	(6)
Predicted probability	0.0281	. 0300			0.8127	0.4345
Constant			3.034***	1.772***		
			(0.0428)	(0.0871)		
P:1	−0.0152					
	(0.00934)					
P:5	−0.0214**	−0.0226**	−0.0127	0.340*	0.183***	0.0368
	(0.00846)	(0.00888)	(0.0536)	(0.182)	(0.0499)	(0.0523)
$P+D$:1	−0.0152					
	(0.00934)					
$P+D$:5	−0.0330***					
	(0.00700)					
$P+D$:10	−0.0211**					
	(0.00849)					
Any msg. but *P*:5		−0.0315**	0.1000**	0.545***	0.210***	0.0765*
		(0.0138)	(0.0468)	(0.127)	(0.0772)	(0.0409)
Quantity ∈ [6,10]			−0.318***			
			(0.0743)			
Quantity ≥ 11			−0.885***			
			(0.0982)			
N	1392	1392	1325	1325	214	1111

Notes: Omitted category is Null message. We estimate Specifications 1, 2, 5, and 6 as probit models and report marginal probabilities. Specifications 1 and 2 use data from all orders, and the others exclude the top 5% of orders. Robust standard errors are in parentheses.
***Indicates 1% significance; **indicates 5% significance; and *indicates 10% significance.

quantile regressions. We jointly estimate the difference between the distribution of Null orders and the distributions of WUSF orders at the 10th, 20th, 30th, 50th, and 75th percentiles. We estimate both the effect of the individual WUSF messages and with all WUSF messages other than *P*:5 grouped together, as in Fig. 6, and we report our results in Table 4. As suggested by the model with fixed preferences in Section 3, the distribution of *P*:5 orders is significantly different from the Null at the 10th and 20th percentiles, which is where consumers may have altered their order sizes under *P*:5 to satisfy the $10 threshold. At the median and above, orders placed under *P*:5 are not significantly different than the Null.[18] The distributions of *P*:1 and $P+D$:δ orders, by contrast, have median values that are significantly different from the Null, as well as significant differences for most of the quantile comparisons we make below the median. Above the median, the point estimates for these messages are generally greater than zero but insignificant due to the variance of order sizes. When all WUSF messages other than *P*:5 are grouped together, their differences from the Null's median and its lower quantiles are consistently positive and significant.

We continue by focusing directly on the $10 threshold for triggering a WUSF donation. In most of this analysis, we condense our results by dividing the treatments into three groups: the Null, the *P*:5 message, and any WUSF message other than *P*:5. We begin with a probit model to estimate the probability that a consumer failed to clear the $10 threshold. Our results, which we report in Specifications 1 and 2 of Table 5, show that consumers with the *P*:5 message were more likely to clear the threshold than consumers who received the Null. We attribute this to consumers with *P*:5 responding to the marginal financial incentives of their WUSF message. [19] The same Specifications also show that consumers with the other WUSF messages were more likely, in general, to exceed the $10 threshold, which may be due to an overall shift in the marginal utility from NPO revenue (w_1). (Similar to the responses of consumers with a *P*:5 message, there were no consumers who received a $P+D$:δ message, donated, and failed to reach the $10 merchandise threshold.) Given the structure of WUSF offers, these results are closely related to Section 3's predictions for choices at the extensive margin.[20]

Next, we examine *how* consumers who received WUSF messages increased their orders' values. In Table 5's Specifications 3 and 4 we show that increased order revenue under *P*:5 was generally due to consumers adding items to their order rather than increasing the average price of items selected. When consumers received a WUSF message other than *P*:5, the average item price increased significantly but by a relatively small magnitude. We include controls in Specification 3 for the quantity of items because consumers with the very largest orders tended to include a large number of small items. The quantity dummy variables in Specification 3 apply to about 10% of all orders in total. In Specification 4 we show that consumers who received the *P*:5 message purchased 0.34 more items, on average, than consumers with the Null, while consumers with other WUSF messages added about half of an item.

We conclude this analysis under the assumption, supported by Specifications 3 and 4, that consumers with *P*:5 messages added items to their orders rather than substituting to higher-priced items. Specifically, we divide consumers into two groups based on whether their final order contains an item priced below $10. Consumers who were not interested in items priced below $10 had no chance to fall below the threshold for a WUSF donation. Consumers with one or more items priced below $10 may have purchased only a single inexpensive item in the absence of a WUSF pledge. For each consumer, we create a new indicator variable for whether her order includes at least two items. This allows us to focus on the quantity margin relevant for consumers deliberately clearing the $10 threshold — virtually all consumers with two or more items spent at least $10. In Table 5's Specification 5 we show that, conditional on their orders containing at least one inexpensive item, consumers who received the *P*:5 message were significantly more likely to purchase at least two

[18] Throughout our analysis, the estimated impact of consumers' responses to treatment messages is robust to whether we add shipping charges to consumers' order values. When we sum merchandise and shipping charges, virtually all orders are above $10. Taken together with the results in Table 4, this is consistent with consumers (correctly) interpreting the WUSF $10 threshold as pertaining to merchandise only, or else consumers with *P*:5 messages could satisfy the WUSF offer terms without altering their order sizes.

[19] Recall from Table 1 that store visitors who received the *P*:5 message purchased with greater probability than recipients of the Null and most other WUSF messages, which is also predicted by the model with stable preferences.

[20] When we estimate the probability that a store visitor placed an order greater than $10, the effect of each WUSF message is positive but insignificantly different from zero.

Table 6
Price sensitivity results.

Dependent variable	Probability of purchase		Probability of purchase		Revenue per visitor		Revenue per visitor	
Included transactions	Any item		Discounted items		Any item		Discounted items	
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted probability	0.020	0.020	0.012	0.012				
Constant					0.900***	0.798***	0.564***	0.410***
					(0.0461)	(0.0956)	(0.0440)	(0.0675)
Discount	0.00100	−0.00156	0.00209**	0.00159	0.225***	0.127	0.215***	0.172
	(0.00106)	(0.00263)	(0.00081)	(0.00204)	(0.0760)	(0.155)	(0.0739)	(0.114)
P:5 message		−0.000357		0.00198		0.0438		0.224
		(0.00255)		(0.00220)		(0.140)		(0.145)
Any msg. but *P*:5		−0.00154		−7.48e-05		0.141		0.174**
		(0.00210)		(0.00165)		(0.113)		(0.0871)
Discount × *P*:5 message		0.00256		−0.00204		−0.0401		−0.317*
		(0.00400)		(0.00246)		(0.215)		(0.188)
Discount × any msg. but *P*:5		0.00329		0.00133		0.158		0.144
		(0.00308)		(0.00234)		(0.185)		(0.152)
N	69,461	69,461	69,474	69,474	69,396	69,396	69,441	69,441

Notes: Omitted categories are Null message and No Discount, as appropriate. We estimate Specifications 1–4 as probit models and report marginal probabilities; these Specifications use all store visits and orders, including the top 5% of orders. We estimate Specifications 5–9 with OLS and exclude the top 5% of orders. Robust standard errors are in parentheses. There are some differences in *N* due to the aggregation of multiple orders by a single consumer; we did not perform this aggregation on orders (or store visits) that resulted in consumers buying only never-discounted items.
***Indicates 1% significance; **indicates 5% significance; and *indicates 10% significance.

items than consumers who received the Null. For consumers with no items priced below $10, we show in Specification 6 that consumers who received the *P*:5 message are not significantly different from the Null in the probability of purchasing two or more items. This is consistent with *P*:5 consumers noticing that their interest in an item priced above $10 is sufficient to generate the WUSF donation, and then proceeding to the check-out without adding any other items.

5.2. Price sensitivity

A goal of the experiment is to determine whether the presence of social messages influences price sensitivity.[21] We analyze this issue by testing whether WUSF pledges affected consumers' responses to discounted prices. In order to have our analysis reflect changes in quantities purchased while also accounting for the variety of items offered at the online store, we make an adjustment to the measurement of consumers' choices. Rather than continuing to look at the total (dollar) spending of each consumer, we examine the amount that a consumer *would have spent* if he faced the undiscounted prices for the choices he actually made. This allows us to evaluate whether a consumer bought "more" in dollar-weighted quantity units when he saw lower prices, and thus permits calculation of demand elasticity. Our analysis is at the session level to account for responses on both the extensive and intensive margins.

In Table 6 we report the results from our analysis of price variation. In Specifications 1, 2, 5, and 6 we include transactions on all items in the store. In the remaining specifications we focus on choices over items that were included in the experiment's pricing component. This includes about half of all product selections and transactions made during the experiment. Consumers who purchased only never-discounted items are treated in the same way as consumers who visited the store and purchased nothing. Similarly, for consumers who selected some never-discounted and some sometimes-discounted items, we include only the latter items. When we analyze the effect of WUSF messages, we divide non-Null treatments by whether the message is *P*:5, as in Table 5.

In the left half of Table 6 we examine purchase probabilities at the session level, which we estimate with a probit model. The analysis of all transactions, in Specifications 1 and 2, reveals few useful results. The standard errors are large relative to the parameter estimates, and none of the coefficients are significantly different from zero. Specification 1 shows a positive but insignificant increase in store visitors' probability of placing an order, and in Specification 2 we find that interactions between discounting and WUSF message indicators yield no significant effects on purchase probability. The results on sometimes-discounted items, in Specifications 3 and 4, are more informative. In Specification 3 we report that consumers who observed lower prices were significantly more likely to order a potentially-discounted item. The coefficients on the interaction terms of Specification 4 are statistically insignificant, suggesting that WUSF messages had no impact on price sensitivity, regardless of whether the message is *P*:5.

In the right half of Table 6 we examine the impact of discounts on total spending by session, using undiscounted prices to calculate price-weighted units selected by consumers. We find that when all sales are grouped together regardless of WUSF message (Specification 5), discounts have a significant and positive effect, increasing order size by about 25% per store visitor. The interaction coefficients in Specification 6 indicate that WUSF messages have no significant effect on price sensitivity. Our estimates in Specifications 7 and 8 are generally more precise, as expected. The overall measure of price sensitivity in Specification 7 is positive and significant at $p<0.01$. This increase of 38% in quantity, when paired with an average discount size of 11.4%, implies a demand elasticity of −3.3 for items at the NPO store. While we find a marginally significant decrease in price sensitivity associated with the *P*:5 message, this result is sensitive to small differences in how we choose to eliminate observations of very large orders.[22]

Although the average effects of price discounts, as measured in Table 6, uncover few differences in demand elasticity with and without WUSF messages, these results may be driven by the price sensitivity of a few large customers.[23] We now turn to an informal analysis of how price variation affected the median consumer within an experimental treatment. We focus on choices over products that

[21] Hiscox and Smyth (2008) study consumers' responses to price changes for products with fair trade labels, and they find that higher prices lead to greater quantity demanded. They attribute this result to consumers giving more credence to fair trade labels when prices are high.

[22] Hiscox and Smyth (2008) study consumers' responses to price changes for products with fair trade labels, and they find that higher prices lead to greater quantity demanded. They attribute this result to consumers giving more credence to fair trade labels when prices are high.

[23] In contrast to the donation pledges of the WUSF messages, the discounting portion of this experiment shifted financial incentives for both large and small purchases.

Table 7
Post-experiment activity.

Dependent variable	Probability of later order				Revenue from later orders			
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted probability	0.0334	0.0335	0.0323	0.0325				
Constant					0.690*	1.269***	0.690*	1.269***
					(0.362)	(0.345)	(0.362)	(0.345)
Supplemental period	0.0130		0.0128		0.760		0.760	
	(0.0183)		(0.0180)		(0.569)		(0.569)	
Any message	0.0243	0.0149*			3.445***	2.866***		
	(0.0150)	(0.00791)			(1.084)	(1.078)		
P:5 message			0.00474	−0.00483			2.646	2.067
			(0.0217)	(0.0135)			(1.810)	(1.806)
Any message but *P*:5			0.0301*	0.0196**			3.655***	3.076**
			(0.0175)	(0.00889)			(1.256)	(1.251)
Null combined with supplemental period?	N	Y	N	Y	N	Y	N	Y
N	2013	2013	2013	2013	2013	2013	2013	2013

Notes: Omitted category is Null message or Null plus Supplement, as appropriate. We estimate Specifications 1–4 as probit models and report marginal probabilities. We estimate Specifications 5–8 with OLS. Robust standard errors are in parentheses.
***Indicates 1% significance; **indicates 5% significance; and *indicates 10% significance.

were included in the discounting portion of the experiment, and we consider pre-discounted order values rather than actual revenue.

There were 57 consumers who purchased potentially-discounted items after receiving the Null and observing undiscounted prices. The median order value is $29.85 for these consumers. Consumers who purchased with discounted prices and the Null message were more numerous (64) and had a greater median order value ($38.93). If each consumer who purchased under the Null was to place an order equal to the median value from his or her respective group, then the total order value would increase by 46% with discounting. Likewise, the 69 consumers who purchased under *P*:5 without discounted prices have a median order of $22, while the 64 consumers with a *P*:5 message and discounts placed a median order of $34.95, which implies that order value would increase by 47% if all orders were at the median. By contrast, the sensitivity at the median was lower for consumers who received a WUSF message other than *P*:5. Regardless of whether they received discounted prices, these consumers placed orders with a median value of $39.95, and the increase in the number of ordering consumers with discounts (from 231 to 288) would yield an increase of 25% in order value if these consumers all purchased at the median. In contrast to the results in Table 6, this suggests that charitable messages may reduce price sensitivity among households making relatively small purchases, provided they cause consumers to update their preferences on a charity's quality.

5.3. Long-term effects of the experiment

While the WUSF messages led to a substantial increase in store revenue during the experiment, a potential concern is that this increased activity came at the expense of other support that consumers may offer the NPO. Consumers could substitute intertemporally and reduce future spending at the store after they purchase under a WUSF message, as occurred in Meier's (2007a) study of stimulating direct donations. Consumers may have also reduced direct donations to the NPO because they see NPO merchandise and donations as substitutes.[24] In terms of the model in Section 3, substitution across periods appears more likely if a WUSF message stimulated a temporary demand response under stable preferences. We address this issue by examining consumers' behavior in the year that followed our experimental messages. While it would be ideal to analyze all consumers who viewed an experimental message while visiting the online store, we are limited to consumers who purchased from the store. Our examination of consumers who purchased is aided by the minimal impact of the WUSF messages on order probabilities, so each observed group contains roughly the same percentage of treated consumers (i.e. those who initiated sessions at the store).

Of the 1325 consumers who purchased under the experiment and were in the lower 95% of orders (divided by type of message or Null), 48 returned to the store in the following year and placed another order. Of these 48, only 4 received the Null message during the experiment. These 4 returning consumers come from a group of 215 who purchased under the Null and were in the lower 95% of orders. Once we add the lower 95% of the supplemental group described in Section 4, we observe a total of 67 consumers who made follow-up purchases from the store, and a total of 2013 consumers from March and April to consider as potentially-returning customers.

We test whether these various groups of consumers behaved differently in their later activity at the store, and we report our results in Table 7. We begin by examining the probability of returning to purchase. In Specification 1 we separate consumers who are in the Null and supplemental groups, and we group together all consumers who received a WUSF message. We find that the supplemental group returned at a slightly higher rate than the Null group, but this difference is not significantly different from zero. By contrast, the consumers who purchased with a WUSF message were nearly twice as likely to return to the store and purchase at least once. In Specification 2 we combine the Null and supplemental groups and again find that the consumers who saw a WUSF message were more likely to return to the store and purchase again. In Specifications 3 and 4 we divide consumers with WUSF messages by whether they received the *P*:5 version. We find that the consumers who we describe above as having shifted preferences returned to the store at a significantly higher rate, while those with *P*:5 messages did not differ significantly from the Null group. Any inference on the *P*:5 message's impact is inhibited, however, by the relatively small number of consumers in a single message category.

In the remainder of Table 7 we examine the revenue from consumers who returned to the store. The dependent variable is the sum of an individual's spending during all return visits, and we include the $0 outcomes of consumers who chose not to return. While we use OLS to minimize distributional assumptions, Tobit analysis of these data provides very similar results with no changes to the statistical significance of any coefficients. In Specifications 5 and 6 we report that consumers who received a WUSF message spent significantly more than those who purchased under the Null or in the supplemental period. In Specifications 7 and 8 we again divide consumers with WUSF messages by whether they received *P*:5, and the results reinforce those on the probability of returning to the store. Consumers who received a non-Null message other than *P*:5 spent

[24] Duncan (1999) explores the closely related topic of consumer substitution between donations of time and money, and he uncovers some intrapersonal crowding-out between the two activities.

significantly more at the store following the experiment, as we might expect if their utility benefit from store purchases, w_1, increased due to exposure to the WUSF message. While consumers who received the *P*:5 message spent more, on average, than consumers in the Null or supplemental group, the differences are not statistically significant.

6. Conclusions

In our field experiment, we find that consumers responded strongly and positively when their merchandise choices could generate revenue for social causes or public goods. Consumers' responses to these opportunities, however, depended on the details of how the revenue was generated. We find that consumers' demand responses were strongest when the explicit financial benefit to the NPO was small (in *P*:1) or even zero (in $P+D$:δ) when consumers declined to make a personal donation required to trigger the benefit. When the NPO's additional financial benefit was relatively large and simple to trigger (in *P*:5), demand increased only near the threshold value for triggering a donation; otherwise the response was no different than the Null. The difference in responses appears to be due to consumers becoming sensitive to financial incentives only when doing so was justified by the financial stakes and the consumer's likely actions. When the financial incentives were not salient, consumers instead changed their behavior in a way consistent with updated preferences or information about the NPO cause.

The unexpected ordering of consumers' responses echoes some of the central results from the literature on intrinsic and extrinsic motivation.[25] This literature includes evidence that offering an agent financial compensation may destroy the agent's willingness to do a task voluntarily. For examples, see Frey and Oberholzer-Gee (1997) or Gneezy and Rustichini (2000). In our study, when sufficient financial incentives draw a consumer's attention, we do not observe the stronger and lower-cost (to the NPO or WUSF) impact of donation pledges on preferences. A difficulty in generalizing from our results is that the difference in consumer response appears to depend on the *level* rather than the *presence* of financial incentives. With our present data, we cannot describe what level of financial incentives is critical for this shift, or whether there is a middle ground where an organization could benefit from both (or neither) type of consumer response.

Our results are also related to recent findings on direct donation decisions, in which the choices of one individual may be sensitive to the nature and extent of other consumers' actions in supporting the same public good. The results in Eckel and Grossman (2003), Karlan and List (2007), Shang and Croson (2009), and Landry et al. (forthcoming), described in the Introduction, are consistent with consumers being aware that they hold imperfect information about a charitable organization's or cause's quality, and positive information on others' support leads consumers to update their preferences for supporting the cause. The WUSF pledges in our study may have the same effect. The presence of a $P+D$:1 pledge, for example, may be relevant to consumers only in that it indicates that some other individuals thought it was worthwhile to create such an offer, which increases the utility value of supporting the NPO.

For for-profit firms considering associations with social causes, our results are clear but perhaps dispiriting for the beneficiary charities. A small donation pledge can stimulate demand more than a large one, and a vague promise of good deeds can be more effective than a specific pledge. These prescriptions, however, depend on how consumers interpret messages about charitable actions. This is an important area for future research. In addition, studies of longer duration and of greater scope would provide useful data on whether the demand stimulation reported here can be replicated as part of a long-term strategy for a firm or nonprofit. While we found only positive or neutral long-run effects from our experimental treatments, our results came from the activity of a relatively small number of consumers who were exposed to a brief demand stimulus.

Acknowledgments

We thank the editor John List, two anonymous referees, Amar Cheema, Dan Elfenbein, Amanda Friedenberg, Gautam Gowrisankaran, Bart Hamilton, Mark Isaac, and Ambar Rao for helpful comments, as well as seminar participants at Virginia, Duke, and the Middlebury College Conference on Philanthropy. We also thank the anonymous nonprofit organization which generously provided the experimental setting and data. This research was supported by grants from the Kauffman Foundation and the NET Institute.

References

Bonini, Mendonca, Oppenheim, 2006. When social issues become strategic, 2. The McKinsey Quarterly, pp. 20–32.
Chen, Yan, Li, Xin, MacKie-Mason, Jeffrey K., 2006. Online fund-raising mechanisms: a field experiment. Contributions to Economic Analysis & Policy, Vol. 5 ((2), Article 4).
Duncan, Brian, 1999. Modeling charitable contributions of time and money. Journal of Public Economics 72 (2), 213–242.
Eckel, Catherine C., Grossman, Philip J., 2003. Rebates versus matching: does how we subsidize charitable contributions matter? Journal of Public Economics 87, 681–701.
Elfenbein, Daniel, McManus, Brian, 2010. A greater price for a greater good? Evidence that consumers pay more for charity-linked products. American Economic Journal: Economic Policy 2, 28–60.
Elfenbein, Daniel, Raymond Fisman and Brian McManus (2010): "Charity as a Substitute for Reputation: Evidence from an Online Marketplace", NBER working paper #15614.
Frey, Bruno S., Oberholzer-Gee, Felix, 1997. The cost of price incentives: an empirical analysis of motivation crowding-out. American Economic Review 87 (4), 746–755.
Gneezy, Uri, Rustichini, Aldo, 2000. Pay enough or don't pay at all. Quarterly Journal of Economics 115 (3), 791–810.
Hiscox, Michael J. and Nicholas F.B. Smyth (2008): "Is There Consumer Demand for Improved Labor Standards? Evidence from Field Experiments in Social Labeling". Harvard University working paper.
Karlan, Dean, List, John A., 2007. Does price matter in charitable giving? Evidence from a large-scale natural field experiment. American Economic Review 97 (5), 1774–1793.
Landry, Craig E., Andreas Lange, John A. List, Michael K. Price, and Nicholas G. Rupp, forthcoming. "Is a Donor in Hand Better Than Two in the Bush? Evidence from a Natural Field Experiment", American Economic Review.
Meier, Stephan, 2007a. 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.
Meier, Stephan, 2007b. A survey of economic theories and field evidence on pro-social behavior. In: Frey, Bruno S., Stutzer, Alois (Eds.), Economics and Psychology: A Promising New Field. MIT Press, Cambridge, pp. 51–88.
Olsen, Douglas G., Pracejus, John W., Brown, Normal R., 2003. When profit equals price: consumer confusion about donation amounts in cause-related marketing. Journal of Public Policy and Marketing 22 (2), 170–180.
Popkowski Leszczyc, Peter T.L., Rothkopf, Michael H., 2010. Charitable intent and bidding in charity auctions. Management Science 56 (3), 399–413.
Sen, Sankar, Bhattacharya, C.B., 2001. Does doing good always lead to doing better? Consumer responses to corporate social responsibility. Journal of Marketing Research 38, 225–243.
Shang, Jen, Croson, Rachel, 2009. Field experiments in charitable contribution: the impact of social influence on the voluntary provision of public goods. The Economics Journal 119, 1422–1439.
Strahilevitz, Michal, Myers, John G., 1998. Donations to charity as purchase incentives: how well they work may depend on what you are trying to sell. Journal of Consumer Research 24, 434–446.

[25] See Meier (2007b) for a review of this literature.