Using Charity Performance Metrics as an Excuse Not to Give

Christine L. Exley *

December 30, 2016

Abstract

There is an increasing pressure to “give wisely.” In a series of experiments, this paper indeed confirms that individuals give less to charities with lower performance metrics. Part of this reduction in giving, however, appears to be driven by individuals overweighting their dislike of low performance metrics as an excuse not to give. Excuse-driven responses to performance metrics persist even when steps are taken to increase the associated cognitive dissonance and are more likely among individuals who give less in general. These results suggest that charities need to carefully consider how and with whom to share performance information.

Keywords: charitable giving; prosocial behavior; altruism; excuses, self-serving biases

*Exley: Harvard Business School, clexley@hbs.edu. Acknowledgements: I gratefully acknowledge funding for this study from the NSF (SES # 1159032).
Sources ranging from Ted talks to third party charity evaluators encourage individuals to give wisely.¹ Charity Navigator, for instance, warns against high overhead costs: “Savvy donors know that the financial health of a charity is a strong indicator of the charity’s programmatic performance [...] the most efficient charities spend 75% or more of their budget on their programs and services.”² GiveWell more generally encourages caution: “The wrong donation can accomplish nothing. Charities that demonstrably change lives are the exception, not the rule.”³

The literature echoes this push towards a focus on charity performance metrics and the professionalization of the nonprofit sector (Hwang and Powell, 2009; van Iwaarden et al., 2009; Ebrahim and Rangan, 2010). Giving decisions are influenced by the benefit size (Eckel and Grossman, 2003; Karlan and List, 2007; Meier, 2007; Eckel and Grossman, 2008; Karlan et al., 2011), whether recipients can choose how to use their donations (Batista et al., 2015), associated overhead costs (Gneezy et al., 2014; Meer, 2014; Metzger and Günther, 2015), measures of effectiveness (Karlan and Wood, 2014), and categorical ratings such as “stars” or endorsements by third party organizations (Gordon et al., 2009; Brown et al., 2016; Yörük, 2016).

While performance metrics clearly influence giving decisions, several findings bring into question the extent to which potential donors inherently value such metrics. First, individuals are often unwilling to pay or to invest effort in learning charity performance metrics.⁴ Among the high-income individuals surveyed in Hope Consulting (2010), 85% state that charity performance is very important but only 35% spend any time researching the performance of charities. In controlled experiments, Null (2011) and Metzger and Günther (2015) document a widespread unwillingness to pay to learn information about the impact of potential donations.⁵

Second, individuals’ unwillingness to support charities that “misuse” donations is not always reflective of a desire for alternative uses. Eckel et al. (2016) find that allowing alumni to direct their donations towards their own academic college increases donations even though the directing option is rarely used. Gneezy et al. (2014) find reduced giving in response to higher overhead costs. They also find, however, that it is not necessary to reduce overhead costs to encourage more giving; ensuring potential donors that others will cover overhead costs proves effective.

Third, performance information appears to have a perverse impact on giving in some cases. Karlan et al. (2011) and Karlan and Wood (2014) observe that “weaker” charity supporters are less likely to give when they receive positive information about the impact of their donations — in particular, that donations will be matched or that the involved aid program is highly effective.

Motivated by these unusual patterns of behavior, this paper proposes a potentially important

---

¹See http://blog.ted.com/2013/03/11/how-to-pick-the-charity-thats-right-for-you
²See http://www.charitynavigator.org/index.cfm?bay=content.view&cpid=419#.U4-blpSwIXo
³See http://www.givewell.org/giving101
⁴Niehaus (2014) models one explanation for this behavior: learning performance information may prevent individuals from maximizing their warm glow by holding the most optimistic beliefs about their impact.
⁵Similarly, Fong and Oberholzer-Gee (2011) observe the most individuals are unwilling to learn whether the recipient of a potential donation is disabled or a drug user, even though they are less willing to give to drug users.
driver of responses to charity performance metrics: they may serve as excuses not to give.\textsuperscript{6} Similar to the excuse-driven responses to risk observed in Exley (2016), individuals may overweight their dislike of low performance metrics as an excuses to keep money for themselves.\textsuperscript{7} To investigate this possibility, this paper conducts three experiments. Participants choose how much to give to low-rated charities in two contexts that only vary in terms of what occurs when money is not given to the low-rated charities.

In the \textit{no tradeoff context}, not giving money to a low-rated charity results in a top-rated charity receiving money. Since participants cannot keep money for themselves (there is no tradeoff between money for themselves and money for charity), excuses are not relevant. The \textit{no tradeoff context} captures non-excuse-driven responses to charity performance metrics.

In the \textit{tradeoff context}, not giving money to a low-rated charity results in participants receiving money. Excuse-driven motives are relevant as participants who respond more negatively to performance metrics receive more money for themselves. Differences in responses to performances metrics across the two contexts are not merely driven participants valuing money for themselves differently than money for a top-rated charity, however. The amount of money participants may choose for themselves in the \textit{tradeoff context} is normalized to be similarly valued to the amount of money they may instead choose for a top-rated charity in the \textit{no tradeoff context}.

The results from all three experiments are consistent with the use of performance metrics as an excuse not to give. While participants give less to charities with lower performance metrics in both contexts, the extent of this reduction is greater in the \textit{tradeoff context} than the \textit{no tradeoff context}. That is, participants appear to overweight the extent to which they dislike lower performance metrics as an excuse to choose money for themselves more often. This excuse-driven behavior is more likely among individuals who also give less to a top-rated charity – i.e., even when lower metrics are not an excuse.\textsuperscript{8}

The first study documents evidence for excuse-driven behavior among laboratory participants. The second study confirms the robustness of the results to a larger pool of online participants. The third study finds that providing an opportunity to prevent donations from being given to a low-rated charity is not sufficient to counter excuse-driven behavior.

Sections 1 - 3 detail the design and corresponding results from the three studies. Section 4 importantly discusses how excuse-driven responses to charity performance metrics may explain the questionable value individuals place on such metrics and thus may help to guide policy.

\textsuperscript{6}Learning about better performance metrics may also result in individuals thinking they need to give less in order to signal prosocial tendencies (Butera and Horn, 2014).

\textsuperscript{7}Related surveys include Merritt et al. (2010); Gino et al. (2016); Bénabou and Tirole (2016). Recent work includes Di Tella et al. (2015); Danilov and Saccardo (2016); Exley and Kessler (2016); Schwardman and van der Weele (2016).

\textsuperscript{8}In addition to potentially relating to the heterogeneous results in Karlan et al. (2011) and Karlan and Wood (2014), similar heterogeneity in excuse-driven behavior is observed in Exley (2016), Exley and Petrie (2016), and Exley and Kessler (2016).
1 Study 1: A Laboratory Experiment

Study 1 involves data from 50 Stanford University undergraduate students in April 2014 who made a series of binary decisions between money for charities and/or money for themselves. In addition to receiving a $20 completion fee, participants knew that one of their decisions would be randomly selected to count for payment. The design and results for Study 1 are as follows.

Three types of charities are involved in Study 1. The first charity type involves three Make-A-Wish Foundation state chapters that vary according to their program expense rates, or percentages of their budgets spent directly on their programs and services (i.e., not spent on overhead costs). In order of their “efficiency” ratings, these state chapters are as follows: the New Hampshire chapter (90%), the Rhode Island chapter (80%), and the Maine chapter (71%). The second charity type involves three Knowledge is Power Program (KIPP) charter schools that vary according to college matriculation rates among their students who completed the 8th grade. In order of their “effectiveness” ratings, these KIPP charter school locations are as follows: Chicago (92%), Philadelphia (74%), and Denver (61%). The third charity type involves three Bay Area animal shelters that vary according to their live release rates, or the combined rates at which animals are adopted, transferred to another rescue organization, or returned to their owner. According to their “effectiveness” ratings, these animal shelters are as follows: the San Francisco SPCA (97%), the Humane Society of Silicon Valley (82%), and the San Jose Animal Care and Services (66%).

So that later decisions in the study account for how participants value money for themselves relative to money for charity, the study begins with the normalization procedure implemented in Exley (2016). In particular, participants begin by completing three “normalization” price lists, one for each charity type. The order in which each participant answers the three normalization price lists is randomly determined, and immediately following each normalization price list, participants complete a “buffer” price list. On each row of a normalization price list for charity \( t \), they choose between (i) $10 for themselves and (ii) some amount for the top-rated charity \( t \) that increases by $2 from $0 to $40 as one proceeds down the twenty-one rows of the price list. The amounts at which they switch to choosing money for the top-rated charity \( t \) imply their \( X_t \) values such that they are indifferent between themselves receiving $10 and the top-rated charity \( t \) receiving \( \$X_t \). Their \( X_t \) values determine the stakes, unbeknownst to participants, involved

---

9See http://www.charitynavigator.org for information on program expense rates.

10See http://www.kipp.org for information on college matriculation rates.


12Buffer price lists involve participants receiving $5 (instead of $10) but are otherwise identical to the normalization price lists. They are intended to limit “stickiness” in participants’ normalization price list decisions across charity types, but unlike the normalization price lists, they do not influence later experimental parameters.

13While even the top-rated charities do not have perfect performance metrics, participants are provided with background information that describes how the top-rated charities exceed a common metric. For Make-A-Wish Foundation state chapters, participants are informed that the highest program expense rate among any state
in the subsequent “valuation” price lists.

Participants complete the valuation price lists in two contexts for each P-rated (i.e., 2nd-rated or 3rd-rated) charity type $t$. The order of valuation price lists – in terms of which charity type $t$ is involved – is randomly determined for each participant. It is also randomly determined whether participants first begin with the no tradeoff context valuation price lists or with the tradeoff context valuation price lists. In both contexts, participants choose between (i) $X_t$ for a P-rated charity $t$ and (ii) some amount that increases by a particular increment as one proceeds down the twenty-one rows of a valuation price list. In the no tradeoff context, this increment involves $\frac{X_t}{20}$ such that the amount the top-rated charity $t$ receives increases from $0$ to $X_t$ as one proceeds down the twenty-one rows of a valuation price list. In the tradeoff context, this increment involves $0.50$ such that the amount participants receive increases from $0$ to $10$ as one proceeds down the twenty-one rows of a valuation price list. Given participants are indifferent between $X_t$ for the top-rated charity $t$ and $10$ for themselves, the row at which participants switch from (i) to (ii) should be the same in both contexts if they are not excuse-driven.\(^{14}\) If participants are excuse-driven, however, they may switch more quickly to (ii) in the tradeoff context to choose money for themselves more often. That is, participants may respond more negatively to low performance metrics in the tradeoff context as an excuse not to give, resulting in lower valuations of options involving low-rated charities.

Before considering participants’ valuations of the low-rated charities, consider how much they value top-rated charities given their $X_t$ values. First, note that $X_t$ values are unclear for three participants who switch from choosing money for themselves to money for the top-rated charity $t$ more than once on a normalization price list. Second, note that $X_t$ values are assumed to equal the censored value of $40$ in $31\%$ of the normalization price lists where participants never switch to choosing some amount for the top-rated charity $t$.\(^{15}\) In all other cases, $X_t$ values are easily inferred by the amounts at which participants switch from choosing $10$ for themselves to $X_t$ for the top-rated charity $t$.\(^{16}\) These $X_t$ values can then be translated into self to charity exchange rates (SCXR), as shown in Appendix Figure A.1. In particular, the average SCXR of 2.78 implies that participants are on average indifferent between $10$ for themselves and $10 \times 2.78$ for the top-rated charity.\(^{17}\) While the main analysis will focus on decisions involving $X_t$ values that can

---

14 This also assumes linearity, which is made more reasonable by the small stakes involved.
15 There are 141 remaining normalization price lists – three for each of the remaining 47 participants. 31% of participants always choosing money for themselves is comparable to the 42% observed in Exley (2016), and more broadly, Engel (2011)’s meta study finding that 36% of dictators do not give anything to their recipients.
16 In particular, to bias against finding excuse-driven evidence, $X_t$ values are estimated as the upper bound of the indifference range implied by participants’ switch points so that they weakly prefer $X_t$ for the top-rated charity $t$ over $10$ for themselves.
17 The average SCXRs range from 2.73 to 2.87 across the three price lists when defined according to their order or when defined according to the charity type involved. These differences are not statistically significant.
be translated into SCXRs, subsequent analyses confirm the robustness to also including decisions that involve censored $X_t$ values.\footnote{Since stronger evidence for excuse-driven behavior emerges for participants with larger $X_t$ values, initially excluding decisions involving censored $X_t$ values seeks to be conservative. It is not possible to include decisions involving the three participants with unknown $X_t$ values due to multiple switch points because the experimental code outputted unreasonable $X_t$ values for them during the experiment.}

In particular, Table 1 presents results from OLS regressions of individuals’ valuations of the 2nd and 3rd ranked charities. Valuations are calculated as a percentage of $10 in the tradeoff context and as a percentage of $X_t$ in the no tradeoff contexts.\footnote{These valuations are set to equal the midpoint of implied ranges from participants’ switch points on the valuation price lists, unless the point estimate of 0% or 100% is implied from a participant never or always choosing $X_t$ for the P-rated charity $t$, respectively.} Note that this is equivalent to considering the row at which participants switch away from choosing $X_t$ for a P-rated charity $t$ in either context, with earlier rows corresponding with lower valuations. In the 7\% of valuation price lists with multiple switch points, valuations are estimated by following prior literature that only considers the first switch points.\footnote{The results are robust to instead excluding any valuations with multiple switch points.}

If participants use performance metrics as an excuse not to give, their valuations of the 2nd and 3rd ranked charities should be lower in the the tradeoff context than the no tradeoff context. Column 1 of Table 1 confirms that this is the case: the average valuation drops by over 10 percentage points if it is elicited in the tradeoff context. When individuals are choosing between money for themselves and money for the lower-rated charities, they appear to overweight their dislike of the low-rated charities as an excuse to choose money for themselves more often. Column 2 confirms that this excuse-driven behavior is robust to the inclusion of individual fixed effects.

Column 3 shows that the drop in valuations in the tradeoff context is more pronounced among individuals who may be more likely to seek excuses not to give: individuals who are more selfish as indicated by higher self to charity exchange rates (SCXR).

Column 4 shows that there are not significant ordering effects between individuals who first complete the tradeoff context valuation price lists (captured by the indicator, order(t,nt)) versus individuals who first complete the no tradeoff context valuation price lists.

Column 5 considers whether excuse-driven behavior varies across charity types. The coefficient on tradeoff context*KIPP shows that there is not a significant difference in how participants respond to the lower college matriculation rates of KIPP charter schools versus the lower program expense rates of Make-A-Wish Foundation state chapters (the excluded charity type). While the positive coefficient on tradeoff context*animal shelters implies relatively less evidence for excuse-driven responses to live release rates of animal shelters, the evidence is still significant.\footnote{From Column 5 in Table 1, the sum of the coefficients on tradeoff context and tradeoff context*animal shelters is significantly different than zero ($p < 0.05$). Follow-up responses suggest this weaker evidence may be driven by some participants thinking that high live release rates are indicative of not needing help. In this sense, either low or high live release rates may serves as excuses. Since such a possibility can confound the results, Studies 2 and 3 examine more unambiguous metrics and strip away unnecessary contextual details about the involved charities.}
Table 1: Ordinary least squares regressions of 2nd and 3rd ranked charity valuations in Study 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>tradeoff context</strong></td>
<td>-10.75**</td>
<td>-10.75**</td>
<td>-10.75**</td>
<td>-15.20***</td>
<td>-17.43***</td>
<td>-23.18***</td>
</tr>
<tr>
<td></td>
<td>(4.00)</td>
<td>(4.19)</td>
<td>(3.13)</td>
<td>(4.32)</td>
<td>(5.25)</td>
<td>(4.77)</td>
</tr>
<tr>
<td>$(SCXR - \overline{SCXR})$</td>
<td>-8.99***</td>
<td>-9.73***</td>
<td>-9.60***</td>
<td>-9.08***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.94)</td>
<td>(3.21)</td>
<td>(3.30)</td>
<td>(1.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>tradeoff context</strong></td>
<td>-12.74***</td>
<td>-11.78***</td>
<td>-11.75***</td>
<td>-13.93***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>(SCXR - \overline{SCXR})</em></td>
<td>(3.40)</td>
<td>(3.42)</td>
<td>(3.44)</td>
<td>(2.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>order(t,nt)</strong></td>
<td>-7.53</td>
<td>-7.46</td>
<td>-8.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.26)</td>
<td>(7.33)</td>
<td>(5.62)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>tradeoff context</strong></td>
<td>9.81</td>
<td>9.84</td>
<td>5.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*order(t,nt)</td>
<td>(6.06)</td>
<td>(6.13)</td>
<td>(6.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KIPP schools</strong></td>
<td>4.54</td>
<td>5.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.54)</td>
<td>(3.55)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>tradeoff context</strong></td>
<td>0.16</td>
<td>-3.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*KIPP schools</td>
<td>(5.01)</td>
<td>(4.17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>animal shelters</strong></td>
<td>-4.15</td>
<td>-2.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.02)</td>
<td>(2.52)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>tradeoff context</strong></td>
<td>6.77*</td>
<td>7.58**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*animal shelters</td>
<td>(3.95)</td>
<td>(3.54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>72.22***</td>
<td>35.91***</td>
<td>72.22***</td>
<td>75.63***</td>
<td>75.47***</td>
<td>70.47***</td>
</tr>
<tr>
<td></td>
<td>(3.99)</td>
<td>(2.29)</td>
<td>(3.47)</td>
<td>(4.64)</td>
<td>(4.84)</td>
<td>(4.00)</td>
</tr>
<tr>
<td><strong>Rating FE</strong></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Ind FE</strong></td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td><strong>Censored X</strong></td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>388</td>
<td>388</td>
<td>388</td>
<td>388</td>
<td>388</td>
<td>564</td>
</tr>
</tbody>
</table>

*p < 0.10, **p < 0.05, ***p < 0.01. Standard errors are clustered at the individual level and shown in parentheses. The above presents OLS regression results of valuations of the P-ranked charity $t$ receiving $X_t$, where $P \in \{2\text{nd}, 3\text{rd}\}$. Valuations are scaled as percentages of $X_t$ in the no tradeoff context and as percentages of $10$ in the tradeoff context. tradeoff context is an indicator for valuations elicited in the tradeoff context. $(SCXR - \overline{SCXR})$ is an individual’s self to charity exchange rate minus the average self to charity exchange rate. order(t,nt) is an indicator for individuals who first complete valuation price lists in the tradeoff context. KIPP schools and animal shelters are indicators for KIPP charter schools or animal shelters, respectively, where the excluded charity type is Make-A-Wish foundation state chapters. “Rating FE” and “Ind FE” indicate whether charity rating fixed effects and individual fixed effects are included, respectively. “Censored X” indicates whether cases involving censored $X_t$ values are included. When cases with censored $X_t$ values are not included, the data include valuations from 31-35 participants for each charity type. When cases with censored $X_t$ values are included, the data include valuations from 47 participants for each charity type.
Column 6 shows that the results are robust to including charity $t$ valuations among participants with censored $X_t$ values.\footnote{The results in Columns 1 - 6 are also robust to instead considering Tobit regressions.}

## 2 Study 2: An Online Experiment

Study 2 involves data from 200 Amazon Mechanical Turk workers in February 2016 who made a series of binary decisions between money for Make-A-Wish foundation state chapters and/or money for themselves. In addition to receiving a $3 completion fee, participants knew that one of their decisions would be randomly selected to count for payment. Relative to the Study 1, participants in Study 2 faced simplified decisions intended to draw greater attention to the performance metrics. Cognitive dissonance from excuse-driven behavior – that may arise from participants’ altering how they respond to performance metrics according to whether their own money is at stake – may therefore be greater in Study 2 than Study 1. In addition to serving as a replication check on a larger sample, Study 2 serves as a plausibly more conservative test of excuse-driven behavior. The design and results for Study 2 are as follows.

Each participant makes decisions involving five Make-A-Wish Foundation state chapters. As part of simplifying the decisions and to avoid any geographical confounds, participants only learn about the performance metric for a particular state chapter, not the actual state involved. In particular, participants are randomly assigned to either learn the program expense rates or the “overall scores” for the state chapters. The five program expense rates involved in this study are described via the following ranges: 85% or better, 80 - 84%, 75 - 79%, 70 - 74%, and 65 - 69%. The five overall scores – which indicate the financial health, accountability and transparency of the state chapters – are described via the following ranges: 95% or better, 90 - 94%, 85 - 89%, 80 - 84%, and 75 - 79%.\footnote{See \url{http://www.charitynavigator.org} for information on how overall scores are calculated.}

As in Study 1 and Exley (2016), this study begins with a normalization procedure to account for any differences in how participants value money for themselves versus money for charity. In particular, on each row of a normalization price list, participants choose between (i) 200 cents for the top-rated charity and (ii) some amount for themselves that increases by 10 cents from 0 to 200 cents as one proceeds down the twenty-one rows of the price list (the charity $t$ notation is dropped since a participant only evaluates one charity type).\footnote{While the normalization price list in Study 1 holds the amount of money for the participants constant, instead holding the amount of money for the top-rated charity constant in Study 2 helps to prevent censored valuations.} The amounts at which they switch to choosing money for themselves imply $Y$ values such that participants are indifferent between 200 cents for the top-rated charity and $Y$ cents for themselves. Their $Y$ values, unbeknownst to participants, determine the stakes involved in the subsequent “valuation” price lists.

Participants complete the valuation price lists in two contexts for each P-rated (i.e., 2nd-rated,
3rd-rated, 4th-rated, or 5th-rated) charity. It is randomly determined whether participants first begin with the no tradeoff context valuation price lists or with the tradeoff context valuation price lists. All valuation price lists involve “tokens” such that any token for a charity yields 2 cents for that charity and any token for a participant yields \( \frac{1}{100} \) cents for that participant. In the no tradeoff context, participants choose between (i) 100 tokens for the P-rated charity and (ii) some amount of tokens for the top-rated charity that increases by 5 tokens from 0 to 100 as one proceeds down the twenty-one rows of a valuation price list. In the tradeoff context, participants choose between (i) 100 tokens for the P-rated charity and (ii) some amount of tokens for themselves that increases by 5 tokens from 0 to 100 as one proceeds down the twenty-one rows of a valuation price list. Given participants are indifferent between 200 cents for the top-rated charity and \( Y \) cents for themselves, the row at which participants switch from (i) to (ii) should be the same in both contexts if they are not excuse-driven.\(^{25}\) If participants are excuse-driven, however, they may switch more quickly to (ii) in the tradeoff context to choose tokens that benefit themselves more often. That is, participants may respond more negatively to low performance metrics in the tradeoff context as an excuse not to give, resulting in lower token-valuations of options involving low-rated charities.

Before turning to participants’ valuations of the lower-rated charities, consider how much they value the top-rated charity given their \( Y \) values. To begin, note that 5% of participants never switch between money for the top-rated charity and money for themselves and thus have “censored” \( Y \) values, which are assumed to equal 200 cents if they always choose money for the charity or 10 cents if they always choose money for themselves. In all other cases, \( Y \) values are easily inferred by the (unique) amounts at which participants switch from choosing 200 cents for themselves to \( Y \) cents for the top-rated charity.\(^{26}\) These \( Y \) values can then be translated into self to charity exchange rates (SCXR), as shown in Appendix Figure A.2. In particular, the average SCXR of 8.33 implies participants are on average indifferent between \( Y \) cents for themselves and \( Y \times 8.33 \) cents for the top-rated charity.\(^{27}\) While the main analysis will again focus on decisions involving \( Y \) values that can be translated into SCXRs, subsequent analyses confirm the robustness to also including decisions that involve censored \( Y \) values.

In particular, Table 2 presents results from several OLS regressions of individuals’ valuations

\(^{25}\)As in Study 1, this assumes linearity, which is made more reasonable by the small stakes involved.

\(^{26}\)That is, unlike in Study 1 where participants could have multiple switch points, the experimental platform in Study 2 forces participants to only have one switch point on the normalization price list (and later valuation price lists). Also, to bias against finding evidence of excuse-driven behavior, the lower bounds of the indifference ranges implied by normalization price list switch points are chosen (when non-zero) so that participants weakly prefer 200 cents for the top-rated charity over \( Y \) cents for themselves. For the 23% of participants with implied indifference ranges from 0-10 cents, the upper bound of 10 cents is instead chosen.

\(^{27}\)The average SCXR of 8.64 among participants viewing program expense rates is not significantly different than the average SCXR of 8.01 among participants viewing overall scores (\( p = 0.57 \)). Also, note that the the notably larger SCXR in Study 2 than in Study 1 likely reflects the normalization price list procedure in Study 2 allowing SCXRs to range form 0 to 20 while the SCXRs could only range from 0 to 4 in Study 1.
of the 2nd - 5th ranked charities. Valuations are considered in terms of token-valuations, which are equivalent to percentage point changes in valuations.\footnote{These valuations are set to equal the midpoint of implied ranges, or if needed because of censoring, to equal the lowest censored valuation of 0 tokens or the highest censored valuation of 100 tokens.} If participants use performance metrics as an excuse not to give, their valuations of the 2nd-5th ranked charities should be lower in the \textit{tradeoff context} than the \textit{no tradeoff context}. Column 1 of Table 2 confirms that this is the case: the average valuation drops by over 5 percentage points if it is elicited in the \textit{tradeoff context}. Column 2 confirms that excuse-driven behavior is robust to the inclusion of individual fixed effects.

Column 3 again finds that excuse-driven behavior is more likely among individuals who are more selfish when excuses related to low performance metrics are not relevant – that is, among participants with higher self to charity exchange rates (SCXR).

Column 4 finds significant ordering effects in terms of whether participants first complete the \textit{tradeoff context} valuation price lists or first complete the \textit{no tradeoff context} valuation price lists. To begin, consider the participants who first complete the \textit{tradeoff context} valuation price lists (captured by the indicator, order\textit{(t,nt)}). The sum of coefficients on \textit{tradeoff} and \textit{tradeoff context}\textit{* order\textit{(t,nt)}} implies that their valuations of low-rated charities are not significantly different across the two contexts.\footnote{The sum of the coefficients on \textit{tradeoff} and \textit{tradeoff context}\textit{* order\textit{(t,nt)}} is not significantly different than zero ($p = 0.57$)} After first evaluating low-rated charities when self-serving motives are relevant in the \textit{tradeoff context}, they subsequently evaluate low-rated charities in a similar manner when self-serving motives are no longer relevant in the \textit{no tradeoff context}. Consistent with a desire to avoid cognitive dissonance from altering how they evaluate low-rated charities across contexts, the negative coefficient on order\textit{(t,nt)} suggests that they achieve this similarity by engaging in more negative evaluations of low-rated charities in the \textit{no tradeoff context} to match how they previously evaluated low-rated charities in the \textit{tradeoff context}.\footnote{Literature documenting a desire to avoid cognitive dissonance includes Babcock et al. (1995); Konow (2000); Haisley and Weber (2010); Gneezy et al. (2016). For a related survey on a desire for belief consonance, see Golman et al. (2016).}

A different pattern of behavior emerges among participants who first complete the \textit{no tradeoff context} valuation price lists. The coefficient on \textit{tradeoff context} implies their valuations of low-rated charities are 13 percentage points significantly lower in the \textit{tradeoff context}. That is, after first evaluating low-rated charities when self-serving motives are not relevant in the \textit{no tradeoff context}, they are subsequently willing to respond more negatively to low performance metrics when self-serving motives are relevant in the \textit{tradeoff context}. Cognitive dissonance that may arise from altering how one evaluates low-rated charities is not sufficient to prevent excuse-driven behavior. That is, there is only evidence for participants’ adjusting their valuations in a manner to avoid cognitive dissonance when doing so is costless to them – i.e., when corresponding adjustments occur in the \textit{no tradeoff context}.\footnote{These valuations are set to equal the midpoint of implied ranges, or if needed because of censoring, to equal the lowest censored valuation of 0 tokens or the highest censored valuation of 100 tokens.}
Column 5 shows that evidence for excuse-driven behavior does not significantly differ according to whether participants are provided with program expense rates or overall scores.

Column 6 shows that the results are robust to the inclusion of the valuations among participants with censored $Y$ values.\(^{31}\)

### Table 2: Ordinary least squares regressions of 2nd - 5th ranked charity valuations in Study 2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(2.29)</td>
<td>(2.44)</td>
<td>(2.24)</td>
<td>(2.82)</td>
<td>(3.45)</td>
<td>(3.72)</td>
</tr>
<tr>
<td>$(SCXR - SCXR)$</td>
<td>-0.75**</td>
<td>-0.76**</td>
<td>-0.74**</td>
<td>-0.75**</td>
<td>-0.75**</td>
<td>-0.75**</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.30)</td>
<td>(0.30)</td>
<td>(0.30)</td>
<td>(0.29)</td>
<td>(0.29)</td>
</tr>
<tr>
<td><strong>tradeoff context</strong></td>
<td>-0.85**</td>
<td>-0.83**</td>
<td>-0.84**</td>
<td>-1.04***</td>
<td>-1.04***</td>
<td>-1.04***</td>
</tr>
<tr>
<td>$(SCXR - SCXR)$</td>
<td>(0.34)</td>
<td>(0.33)</td>
<td>(0.33)</td>
<td>(0.33)</td>
<td>(0.32)</td>
<td>(0.32)</td>
</tr>
<tr>
<td><strong>order(t,nt)</strong></td>
<td>-8.49**</td>
<td>-8.86**</td>
<td>-8.44**</td>
<td>-8.44**</td>
<td>-8.44**</td>
<td>-8.44**</td>
</tr>
<tr>
<td></td>
<td>(4.13)</td>
<td>(4.11)</td>
<td>(4.11)</td>
<td>(4.11)</td>
<td>(4.02)</td>
<td>(4.02)</td>
</tr>
<tr>
<td><strong>tradeoff context</strong></td>
<td>15.05***</td>
<td>15.21***</td>
<td>14.27***</td>
<td>-2.37</td>
<td>-4.37</td>
<td>-4.37</td>
</tr>
<tr>
<td>$(SCXR - SCXR)$</td>
<td>(4.33)</td>
<td>(4.35)</td>
<td>(4.35)</td>
<td>(4.35)</td>
<td>(4.35)</td>
<td>(4.35)</td>
</tr>
<tr>
<td><strong>overall score</strong></td>
<td>5.41</td>
<td>5.51</td>
<td>5.51</td>
<td>5.51</td>
<td>5.51</td>
<td>5.51</td>
</tr>
<tr>
<td>$(SCXR - SCXR)$</td>
<td>(4.34)</td>
<td>(4.35)</td>
<td>(4.35)</td>
<td>(4.35)</td>
<td>(4.35)</td>
<td>(4.35)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>36.68***</td>
<td>17.97***</td>
<td>36.68***</td>
<td>40.95***</td>
<td>38.47***</td>
<td>38.53***</td>
</tr>
<tr>
<td></td>
<td>(2.23)</td>
<td>(1.50)</td>
<td>(2.22)</td>
<td>(2.85)</td>
<td>(3.44)</td>
<td>(3.41)</td>
</tr>
</tbody>
</table>

* p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the individual level and shown in parentheses. The above presents OLS regression results of valuations of the P-ranked charity receiving 200 cents, where $P \in \{2nd, 3rd, 4th, 5th\}$. Valuations are in tokens, where each token for a charity yields 2 cents and each token for a participant yields $\frac{Y}{100}$ cents. **tradeoff context** is an indicator for valuations elicited in the **tradeoff context**. $(SCXR - SCXR)$ is an individual’s self to charity exchange rate minus the average self to charity exchange rate. **order(t,nt)** is an indicator for individuals who first complete valuation price lists in the **tradeoff context**. **overall score** is an indicator for participants’ decisions involving overall scores instead of program expense rates of Make-A-Wish foundation state chapters. “Rating FE” and “Ind FE” indicate whether charity rating fixed effects and individual fixed effects are included, respectively. “Censored Y” indicates whether individuals with censored $Y$ values are included. When individuals with censored $Y$ values are not included, the data include valuations from 189 participants. When individuals with censored $Y$ values are included, the data include valuations from 200 participants.

\(^{31}\)The results in Columns 1 - 6 are also robust to instead considering Tobit regressions.
3 Study 3: An Online Experiment with A Policy Test

One interpretation of Study 2 is that it documents evidence for a dislike of cognitive dissonance, just not a sufficient dislike to counter excuse-driven behavior. An effective strategy for countering excuse-driven behavior may therefore involve increasing the associated cognitive dissonance. Study 3 investigates this possibility by collecting decisions from an additional 201 Amazon Mechanical Turk workers in April 2016. In addition to receiving a $3 completion fee, these participants knew that one of their decisions would be randomly selected to count for payments. The design and results – that vary according to which group participants were randomly assigned – are as follows.

The “control” group design replicates Study 2’s design with one minor exception: all participants first complete the no tradeoff context valuation price lists and then the tradeoff context valuation price lists. The “intervention” group design only differs from the control group in that, if participants choose for any money to be given to low-rated charities in the tradeoff context, they can choose to redirect this money to the top-rated charity by completing a simple effort task. Successfully completing the effort task involves correctly counting the number of 0s that are in a series of 400 numbers. Participants are aware of this opportunity when deciding whether to keep money for themselves or to give money to charity in the tradeoff context.

Using performance metrics as an excuse not to give may be more difficult for participants in the intervention group regardless of whether they intend to complete the effort task. Participants who intend to complete the effort task may view their charitable giving decisions in the tradeoff context as only involving the top-rated charities. Participants who do not intend to complete the effort task may find it difficult to simultaneously believe that it is not worthwhile to put forth effort to redirect money from a low-rated charity to the top-rated charity and that low performance metrics serve as good excuses not to give.

Table 3 presents results from several OLS regressions of individuals’ valuations of the 2nd - 5th ranked charities. Column 1 confirms excuse-driven behavior persists overall: the average valuation drops by over 14 percentage points if it is elicited in the tradeoff context. As seen by the coefficient on tradeoff context*intervention, however, this drop is not significantly reduced among participants in the intervention group. In line with the persistence of excuse-driven behavior, 61% of participants in the intervention group do not even guess the number of 0s and only 8% of participants provide the correct number.

Column 2 confirms that excuse-driven behavior is again more likely among individuals who are more selfish as indicated by higher self to charity exchange rates (SCXR).\[^{32}\]

Column 3 shows that while participants’ valuations are significantly different if they evaluate

\[^{32}\]Appendix Figure A.3 shows the self to charity exchange rates (SCXR), which on average equal 8.94. The average SCXR of 8.78 among participants viewing program expense rates is not significantly different than the average SCXR of 9.10 among participants viewing overall scores (p = 0.77).
Make-A-Wish foundations that vary according to overall scores, instead of program expense rates, significant evidence for excuse-driven behavior persists in both cases.

Column 4 shows that the results are robust to the inclusion of the valuations among participants with censored Y values.33

Table 3: Ordinary least squares regressions of 2nd-5th ranked charity valuations in Study 3

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>tradeoff context</td>
<td>-14.12***</td>
<td>-14.34***</td>
<td>-10.40***</td>
<td>-10.30***</td>
</tr>
<tr>
<td></td>
<td>(3.16)</td>
<td>(2.99)</td>
<td>(3.50)</td>
<td>(3.49)</td>
</tr>
<tr>
<td>intervention</td>
<td>2.69</td>
<td>2.87</td>
<td>2.05</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>(3.99)</td>
<td>(3.96)</td>
<td>(3.90)</td>
<td>(3.85)</td>
</tr>
<tr>
<td>tradeoff context</td>
<td>2.84</td>
<td>3.29</td>
<td>4.15</td>
<td>5.77</td>
</tr>
<tr>
<td>*intervention</td>
<td>(4.66)</td>
<td>(4.40)</td>
<td>(4.37)</td>
<td>(4.25)</td>
</tr>
<tr>
<td>(SCXR − SCXR)</td>
<td>-0.52*</td>
<td>-0.54*</td>
<td>-0.64**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.29)</td>
<td>(0.28)</td>
<td></td>
</tr>
<tr>
<td>tradeoff context</td>
<td></td>
<td>-1.34***</td>
<td>-1.33***</td>
<td>-1.50***</td>
</tr>
<tr>
<td>*(SCXR − SCXR)</td>
<td></td>
<td>(0.31)</td>
<td>(0.31)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>overall score</td>
<td></td>
<td>8.66**</td>
<td>9.11**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.91)</td>
<td>(3.88)</td>
<td></td>
</tr>
<tr>
<td>tradeoff context</td>
<td></td>
<td></td>
<td>-9.01**</td>
<td>-10.33**</td>
</tr>
<tr>
<td>*overall score</td>
<td></td>
<td></td>
<td>(4.38)</td>
<td>(4.25)</td>
</tr>
<tr>
<td>Constant</td>
<td>38.57***</td>
<td>38.48***</td>
<td>34.69***</td>
<td>35.97***</td>
</tr>
<tr>
<td></td>
<td>(2.79)</td>
<td>(2.78)</td>
<td>(3.20)</td>
<td>(3.18)</td>
</tr>
</tbody>
</table>

* p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the individual level and shown in parentheses. The above presents OLS regression results of valuations of the P-ranked charity receiving 200 cents, where P ∈ {2nd, 3rd, 4th, 5th}. Valuations are in tokens, where each token for a charity yields 2 cents and each token for a participant yields 100 cents. tradeoff context is an indicator for valuations elicited in the tradeoff context. (SCXR − SCXR) is an individual’s self-to-charity exchange rate minus the average self-to-charity exchange rate. intervention is an indicator for individuals in the intervention group. overall score is an indicator for participants’ decisions involving overall scores instead of program expense rates of Make-A-Wish foundation state chapters. “Rating FE” indicate whether charity rating fixed effects are included. Individual fixed effects are not included to allow for examination the intervention impact. “Censored Y” indicates whether individuals with censored Y values are included. When individuals with censored Y values are not included, the data include valuations from 188 participants. When individuals with censored Y values are included, the data include valuations from 201 participants.

33The results in Columns 1 - 4 are also robust to instead considering Tobit regressions.
4 Conclusion

If individuals use charity performance metrics as an excuse not to give, as documented in the three studies in this paper, it may naturally follow that individuals place questionable value on such metrics. Giving less to low-rated charities than to top-rated charities need not indicate strong preferences for better performance metrics. Individuals may merely find it easier to decline the ask when they can use the excuse of lower metrics.

The most cost-effective ways for non-profit organizations to counter such excuse-driven behavior is not straightforward. Eliminating or not providing performance information may prove ineffective, particularly since the lack of performance information may foster the development of excuses. Meanwhile, the returns to investing in improvements to performance metrics may be limited if even the smallest shortcomings can serve as excuses not to give. Improving performance metrics may also come at the cost of hindering an organization’s effectiveness by shortchanging fundraising expenditures or by discouraging innovation.

Future work on how best to develop and to provide charity performance metrics may take into account promising insights from prior literature and this study, however. For instance, the framing of overhead costs in Gneezy et al. (2014) may be effective because it counters the psychological development of excuses: individuals cannot point to “their donation” as not having a direct impact if other donors cover associated overhead costs. The option to direct the use of donations in Eckel et al. (2016), even though individuals do not take advantage of it, may be effective because it eliminates excuses around a desire to have the donations used differently. Adverse responses to positive impact information among “weaker” supporters, such as in Karlan and Wood (2014), may occur because individuals who give less in general are more susceptible to excuses and thus likely to view positive, but imperfect, impact information unfavorably.

The flip side of this heterogeneity — that individuals who give more in general are less susceptible to excuses — is perhaps even more policy relevant. Targeting this group of less-excuse-driven individuals may be crucial to the funding of non-profit organizations desiring to embark on high risk but high returns endeavors, such as developing new vaccines or technological innovations to help those in need. Providing them with performance metrics may also help to guide giving towards the most effective opportunities without fear of the unintended consequence of highlighting excuses not to give.

---

34 For instance, Exley and Petrie (2016) find that individuals develop their own excuses not to give when no impact information is provided. Haisley and Weber (2010) document how individuals develop self-serving views of ambiguity. To achieve more selfish outcomes, the “moral wiggle room” literature (Dana et al., 2007; Bartling et al., 2014; Grossman, 2014; van der Weele, 2014; Grossman and van der Weele, 2016) shows how individuals avoid information and the avoidance of the ask literature (Andreoni et al., 2016; DellaVigna et al., 2012; Trachtman et al., 2015; Lin et al., 2016) shows how individuals sidestep being asked to give altogether.

35 For instance, see Andreoni and Payne (2011) and Grant (2011).

36 Relatively little work focuses on the motivations to give among those who give the most or who are able to give the most. For one related study, however, see Kessler et al. (2016).
References


Andreoni, James, Justin M. Rao, and Hannah Trachtman. 2016. “Avoiding the ask: A field experiment on altruism, empathy, and charitable giving.” *Journal of Political Economy*.


Eckel, Catherine C., David Herberich, and Jonathan Meer. 2016. “A field experiment on directed giving at a public university.” *Journal of Behavioral and Experimental Economics*.


15
Grossman, Zachary. 2014. “Strategic ignorance and the robustness of social preferences.” 

Grossman, Zachary, and Joël J van der Weele. 2016. “Self-image and willful ignorance in 
social decisions.” *Journal of the European Economic Association*.

Haisley, Emily C., and Roberto A. Weber. 2010. “Self-serving interpretations of ambiguity 

Hope Consulting. 2010. “Money for Good The US Market for Impact Investments and Char-
ritable Gifts from Individual Donors and Investors.” *Technical Report*.

ences of Professionalism in the Nonprofit Sector.” *Administrative Science Quarterly*, 54: 268–
298.

effectiveness in a direct mail fundraising experiment.” *NBER Working Paper Series, Working 
Paper 20047*.

1774–1793.

Karlan, Dean, John A List, and Eldar Shafir. 2011. “Small matches and charitable giving: 

and Powerful to Give.” *Working paper*.


Lin, Stephanie C., Rebecca L. Schaumberg, and Taly Reich. 2016. “Sidestepping the 
rock and the hard place: The private avoidance of prosocial requests.” *Journal of Experimental 

Meer, Jonathan. 2014. “Effects of the price of charitable giving: Evidence from an online 

donations in a field experiment.” *Journal of the European Economic Association*, 5(6): 1203–
1222.


A Appendix

Figure A.1: Distribution of SCXR in Study 1

Each bar shows the percent of the participants in Study 1 with a given self to charity exchange rate (SCXR), which equals $\frac{X_t}{10}$ for each charity $t$, where participants are indifferent between $X_t$ for the top-rated charity $t$ and $10$ for themselves. The results include data for the 47 participants with accurately estimated $X_t$ values.

Figure A.2: Distribution of SCXR in Study 2

Each bar shows the percent of the participants in Study 2 with a given self to charity exchange rate (SCXR), which equals $\frac{200}{Y}$, where participants are indifferent between 200 cents for the top-rated charity and $Y$ cents for themselves. The results include data for the 189 participants with accurately estimated $Y$ values.

Figure A.3: Distribution of SCXR in Study 3

Each bar shows the percent of the participants in Study 3 with a given self to charity exchange rate (SCXR), which equals $\frac{200}{Y}$, where participants are indifferent between 200 cents for the top-rated charity and $Y$ cents for themselves. The results include data for the 189 participants with accurately estimated $Y$ values.