Journal of Behavioral and Experimental Economics 103 (2023) 101988

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

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

Demanding the morally demanding: Experimental evidence on the effects of moral arguments and moral demandingness on charitable giving

Ben Grodeck [a,*,#], Philipp Schoenegger [b,#]

[a] Monash University, Monash Business School, Department of Economics
[b] University of St Andrews, School of Economics and Finance & School of Philosophical, Anthropological and Film Studies

ARTICLE INFO

JEL codes:
D64
D91
H41
C90

Keywords:
Charitable giving
Experiment
Morality
Obligation
Pro-social behavior

ABSTRACT

What are the effects of confronting people with moral arguments and morally demanding statements to perform certain actions, such as donating to charity? To investigate this question, we conduct an online randomized experiment via Prolific (n=2500) where participants can donate to charity. Using a between-subject design, we provide some participants with a moral argument as to why they should donate. We then add a single sentence on top of the moral argument that expresses and varies moral demandingness at different levels. To reduce experimenter demand worries, in a follow-up experiment (n=1200) we provide the same stimulus via an external party's website—the non-profit Giving What We Can. In both experiments, we find that moral arguments significantly increase both the frequency and amount of donations compared to the control. However, we fail to find evidence that increasing the level of moral demandingness affects donation behavior in either experiment. Exploratory equivalence tests provide evidence in favor of such a null effect. Our findings suggest that charities should employ moral arguments to increase giving, though our findings give no specific recommendation as to the moral demandingness employed as there is no additive effect of morally demanding arguments.

1. Introduction

Many[1] philosophers argue that we have a moral obligation to donate to charity (e.g., Singer, 1972; MacAskill, 2015; Pummer, 2016; Berkey, 2020). However, they also worry that communicating this obligation may lead to a backfire effect, reducing the desired behavior. This is important because moral arguments that state morally demanding obligations are central to many domains of behavior, among them charitable giving, climate change activism, and social justice movements. They all share the underlying structure in which a fundamental moral claim directly leads to a moral obligation about a certain type of behavior, be it donating a portion of one's income to charity, reducing one's greenhouse gas emissions, becoming vegetarian, or adopting new phrases in everyday speech. A common approach here is to tell people why they should perform these actions and subsequently state how obligated they are to act this way. For example, after providing the argument for an action X one might communicate how morally demanding X is. For instance, one might tell a person that they can do a lot of good by doing X, that they should do X, that they are morally obligated to do X, or that it is immoral for them not to do X. It is far from clear what the optimal communication strategy with regard to these moral demands is. In other words, how should one communicate this moral obligation to perform an action with the aim to elicit the highest possible compliance? Plausibly, not engaging in morally demanding language at all might leave behavior change untapped but drawing on too morally demanding phrases might in turn lead to a backfire effect.

* Corresponding author.
E-mail address: ben.grodeck@gmail.com (B. Grodeck).
# Authors are ordered alphabetically. Both authors contributed equally to this paper.

[1] We want to thank Philip Grossman, Benjamin Sachs, Zach Freitas-Groff, Theron Pummer, Lata Ganghadaran, Erte Xiao, Brian Jabarian, Kirby Nielsen, David Rhys Bernard, Joshua Miller, Hayden Wilkinson, Tyler John, Miguel Costa-Gomes, Peter Singer, Toby Handfield, David Reinstein and Nina Xue for helpful comments and suggestions.This work was supported by a grant from the Forethought Foundation and the Centre for Effective Altruism as well as additional research funding from Giving What We Can.

https://doi.org/10.1016/j.socec.2023.101988
Received 13 June 2022; Received in revised form 17 January 2023; Accepted 30 January 2023
Available online 2 February 2023

These questions have direct implications for communication strategies that employ moral arguments in a wide variety of contexts and as such ought to be studied carefully.

Consider the following specific case study: As Ben Sachs remarks in recent work in the context of charitable giving (Sachs, 2019), even (meta-)charitable organizations aligned with the Effective Altruism movement[2] do not communicate high levels of moral demandingness in their public facing materials. For example, Giving What We Can, a non-profit founded to motivate individuals to donate to highly effective charities, shies away from explicitly moral language across their communication channels at the time of writing. As Sachs speculates, this proclivity for non-moral language might be due to a concern about the counterproductivity of communicating high moral demandingness (Sachs, 2019, 2). Plausibly, the best explanation for the communication strategy observed above is the underlying worry that expressing strong moral demandingness may backfire and lead to lower, not higher, donations. A number of philosophers working on charitable giving have raised this specific concern (e.g., Kagan, 1989; Unger, 1996; De Lazari-Radek and Singer, 2010). They suggest that confronting people with morally demanding requirements might be counterproductive to the overarching goal of maximizing donations. However, this is an empirical question that has not yet been answered directly. In this paper, we aim to provide such an answer.

To do this, we conduct two online experiments on Prolific where participants can donate to the charity GiveDirectly that provides direct cash-transfers to those living in extreme poverty. Using a between-subject design, we randomly assign participants either into a control group, a moral argument treatment, or one of three moral demandingness treatments. We provide the control group with paragraphs about UK parliamentary procedure, while all other groups receive information concerning the situation of global poverty as well as an argument for how they can help those living in extreme poverty due to their relative affluence. In our three moral demandingness treatments, we add to the moral argument text an additional sentence that varies the moral demandingness of the argument across conditions from low to high demandingness.[3] By adding just a single sentence, our experimental design allows us to cleanly identify the effect that incremental increases in moral demandingness have on donation behavior above and beyond the effect of the moral argument.

We also report a follow-up experiment that has two aims: (i) replicate the results from the first experiment and (ii) reduce what we perceived to be the biggest design weakness of the previous experiment: experimenter demand. In this follow-up experiment, we solely focus on differences between the control, the moral argument, and the strong demandingness conditions to replicate the previous results of disparate gender effects in the strong demandingness condition and the effect of the moral argument over the control. The central design improvement over the first experiment is that we provide participants with the treatment text via a third-party website, the organization Giving What We Can.[4] This was done to mitigate possible experimenter demand effects that might arise from us—the experimenters[5]—providing the moral argument and moral demandingness statements to participants.

In both experiments, we find that providing people with moral arguments significantly increases the frequency of donations made by 22.4% and 27.8% and the amount donated by 51.7% and 42.9% respectively compared to the Control. However, adding a single sentence on top of the moral argument, changing the moral demandingness between treatments (ranging from Inspiration with the lowest moral demandingness to Strong Demandingness with the highest moral demandingness) does not increase the frequency or the amount donated in the aggregate. However, in the first experiment, we find a significant gender effect, with strong moral demandingness representing a backfire for males. In our second experiment, we replicate this aggregate result with the Strong Demandingness treatment having no impact on the frequency or amount of donations made and fail to replicate our gender effect. As such, we conclude that we fail to find significant effects of moral demandingness on charitable giving above and beyond the effect of the moral argument.

This present paper is best understood as contributing to the general literature on charitable solicitation[6] and pro-social behavior more broadly. Regarding pro-social behavior, Bursztyn et al. (2019) show that communicating a moral appeal to (in-)justice is associated with reduced delinquency and default rates in debt repayment. This suggests that there is some underlying relationship between moral claims such as those relating to justice and monetary decisions that may generalize outside of their context. Similarly, Dal Bo and Dal Bo (2014) show that moral suasion increases contributions in a series of voluntary contribution game laboratory experiments. However, field experiments that attempt to implement the same mechanism in the context of tax payments show mixed results (Torgler, 2004; Blumenthal et al., 2001). Moral suasion interventions differ from the moral demandingness studied here because most moral suasion interventions, which include statements like 'Don't lie,' 'It is wrong to steal,' or 'Be a team player' do not make explicit the moral demandingness underlying the claim. Specifically manipulating this level of demandingness has not been done before in the way presented here. The moral suasion literature also differs from our moral argument in that suasion often involves statements and proclamations, while we present participants with background information and a moral argument based on the evidence, rather than making a standalone normative claim.

Relating to the effectiveness of moral suasions, Ito et al., (2018) find that economic incentives outperform moral suasion both in the short-term effect sizes and longer-term persistence of effects (cf. also Björn et al., 2020).[7] While some studies have looked at how moral suasion and moral arguments affect behavior, explicitly varying the levels of the moral demandingness has been neglected in the literature. Our study contributes to this literature by investigating how different levels of morally demanding appeals impact charitable giving behavior.

The results from our experiments show the importance of language and communication on charitable giving. Moral arguments—specifically related to relative affluence—are an effective tool in changing donation

[2] Effective Altruism is a social movement focused on using evidence to do the most good given the available resources. This includes researching and donating to the highest cost-effective charitable interventions such as malaria bed nets to researching global catastrophic risks like those posed by artificial intelligence or nuclear war. At the core of its mission is the claim that we all ought to do more than we already are doing, and that our contributions ought to be effective.

[3] We validate the normative differences between the three statements in an ancillary experiment (See Appendix B).

[4] For the control treatment, we used the website that the control paragraph was originally taken from.

[5] One might argue this does not eliminate experiment demand effects completely, as we are still showing GWWC's information, which may be seen as an implicit endorsement. However, it is plausible to argue that the experimenter demand effects are lower in Experiment 2 compared to Experiment 1.

[6] Most of charitable giving happens in response to a solicitation attempt of one kind or another (Bekkers and Wiepking, 2010; Bryant et al., 2003; Bekkers, 2005). Their effectiveness may depend on variables such as frequency (Meer, 2011; Meer and Rosen, 2012), transaction costs (Huck and Rasul, 2010), matching rates (Eckel and Grossman, 2003; Karlan and List, 2007), type of appeals (List et al., 2019), or the amount either directly requested or suggested (Shang and Croson, 2009; Edwards and List, 2014; Adena et al., 2014; Reily and Samek 2019; Altmann et al., 2019; Ekström, 2021).

[7] For further recent usage of moral nudges, see Capraro et al. (2019), Böhm et al. (2020) and Bos et al. (2020).

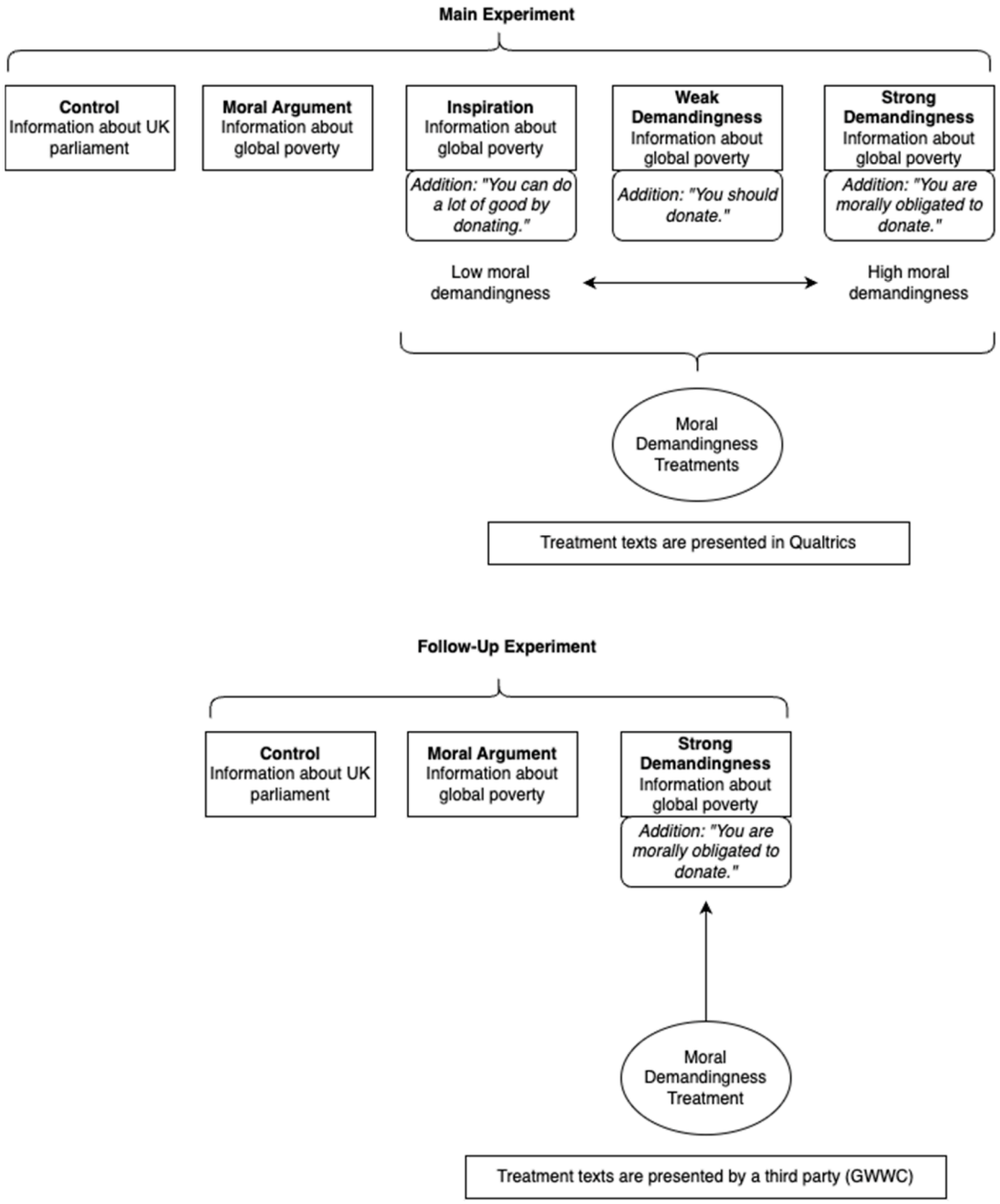

Fig. 1. Experimental Conditions.

behavior. However, there is no robust effect of morally demanding language across both experiments that would go above and beyond the moral arguments. This may imply that we overestimate the effects of this type of language on behavior, or that when the moral argument is powerful enough, it waters down effects of morally demanding communication. This paper opens space for future research on the impact of moral language and communication on charitable giving behavior.

2. Experiment 1

For a graphical overview of all experimental conditions and the additional sentences that distinguish the treatments (in both Experiment 1 and 2), see Fig. 1.

2.1. Participants

For our first experiment, we collected our sample via Prolific, an online participant recruitment platform.[8] We advertised the study as a time-sensitive puzzle-solving task to mask the main purpose to reduce potential self-selection, however, all information portrayed was fully accurate. We informed participants in our study that their task was to

[8] The use of platforms like Prolific (or MTurk) has already received significant uptake in experimental/behavioral economics (e.g., Hauser and Schwarz, 2016; Gandullia and Lezzi, 2018; Palan and Schnitter, 2018); Gandullia, 2019; Giamattei et al., 2020). As Gandullia et al. (2020, 2) have argued, moving from a university student sample to an online sample may also reduce experimenter demand effects as the experimenters are not physically present at the time of data collection thus further justifying this choice of participant recruitment.

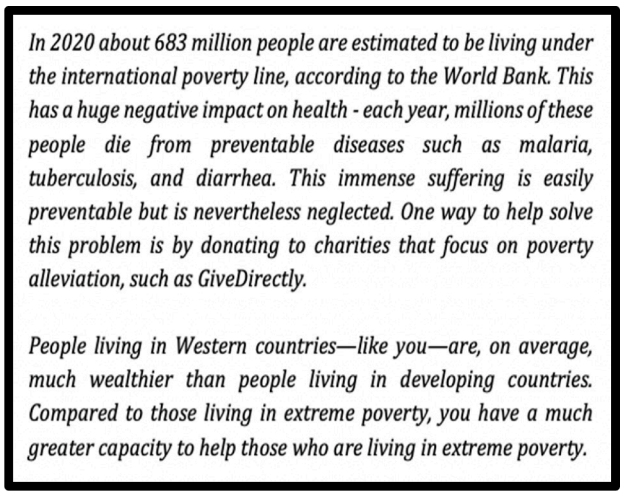

Fig. 2. Moral Argument paragraphs.

solve Raven's matrices. We portrayed the full accurate payment structure on Prolific before participation. We recruited participants exclusively from the United Kingdom and provided a base pay of £1 for participation. In addition, we gave participants the opportunity to earn an additional £1 bonus endowment for completing the Raven's matrices. We recruited 2500 subjects to participate in the study. We used G*Power (Faul et al., 2007) to conduct an a priori power analysis to arrive at this sample size. Our goal was to obtain .80 power to detect an effect size of .18 (Cohen's d) at a standard .05 alpha error probability. This required 486 participants in each group for a two-sided t-test. This was to account for eliminated participants across our five conditions based on the attention check. Overall, 270 participants across all conditions failed the main attention check. We dropped them from the final data set and excluded them from all analyses. Before data collection, we preregistered the study's design and aims on the Open Science Framework.[9,10,11]

2.2. Experimental design

The study consisted of three parts[12]: (1) A real-effort puzzle-solving task, (2) a donation task, (3) and a final questionnaire. The real effort task included six Raven's matrices, the first of which functioned as a demonstration and attention check. We timed all the Raven's matrix tasks and informed participants that they had 30 seconds to complete each problem. After these 30 seconds, the next problem automatically appeared in place of the previous one. We also incentivized participants to complete the matrices as quickly and accurately as possible by offering an additional bonus of £5 for the individual who completed the most problems correctly in the shortest time. All participants received their advertised bonus endowment after completing the matrices task. Throughout the experiment, we portrayed the bonus endowments in experimental currency units (ECUs). We rewarded participants with 20 ECUs for completing the matrices irrespective of performance and informed them that the exchange rate was 1 ECU = £0.05.

In the second part of the experiment, we provided participants with information about GiveDirectly. Then, we randomly assigned participants into one of five conditions (see Fig. 1).[13] In the first condition, 'Control', we presented participants with two paragraphs on parliamentary procedure in the UK. These paragraphs were roughly equal in word count and paragraph structure to all other treatments. This condition functioned as a content control condition against which we could compare to the Moral Argument treatment. Then we asked participants to answer two comprehension questions. If at least one of their answers was incorrect, we once again showed them the paragraphs and asked them to answer the comprehension questions again. If, on this second attempt, they again failed at least one of the comprehension questions, we informed them of the correct answers. They then proceeded in the experiment.

In the second condition, 'Moral Argument', participants received the following two paragraphs constituting the basic moral argument (see Fig. 2). This produces a moral argument as the informational content provides moral reasons to donate to charities in developing countries. This condition functioned as the control condition against which we compare all the moral demandingness treatments that vary the level of demandingness. It also allowed us to observe the effect of moral arguments on donation behavior by comparing it to the control.

The third, fourth, and fifth conditions all included the same paragraphs as the Moral Argument condition. Crucially, all three moral demandingness treatments included an additional bolded sentence varying the moral demandingness statement to act. The extra sentence appeared at the bottom of the text in Fig. 2. Specifically, the difference in the intervention sentences were:

> **Inspiration Treatment:** "For these reasons, you can do a lot of good if you give money to charities–such as Give Directly–to alleviate the suffering of people in developing countries at a minimal cost to yourself"
> **Weak Demandingness Treatment:** "For these reasons, you should give money to charities–such as Give Directly–to alleviate the suffering of people in developing countries at a minimal cost to yourself"
> **Strong Demandingness Treatment:** "For these reasons, you are morally obligated to give money to charities–such as Give Directly–to alleviate the suffering of people in developing countries at a minimal cost to yourself."

The three moral demandingness treatments represent different levels of moral demandingness, ranging from a purely inspiring message to one that explicitly highlights everyone's moral obligation to donate. We ran a pre-experimental study and found that participants independently evaluated these three statements to have different levels of moral demandingness—with Inspiration being the lowest and Strong Demandingness being the highest as outlined in Fig. 1 (see Appendix B for experimental details and results of this pre-experimental study). Fig. 3 outlines the experimental procedure including final sample sizes for each condition.

We also preregistered secondary analyses based on demographics and further variables collected in the last part of the main study. Specifically, these are the effects of moral demandingness on donation behavior conditional on gender, altruistic type, and utilitarian attitudes, as well as self-reported feelings of obligation. We asked participants to complete the Oxford Utilitarian Scale (OUS) (Kahane et al., 2018)[14],

[9] https://osf.io/82mz4/?view_only=61b0a1c22cb54256a621ca335f05ca96.
[10] We report the following deviations from the preregistration: First, we preregistered a base pay of £0.90 but ended up paying £1 to ensure a fairer pay in case participants took longer than anticipated. Second, we designated 'gender' as a variable used for secondary analyses. In this paper we report the gender results alongside the primary results. Third, we stated that we would run ANOVAs to test for group differences. Because of the disciplinary background of this paper, we decided not to do this.
[11] We received ethics approval for this study from the University Teaching and Research Ethics Committee (UTREC) at the University of St Andrews: SA15064.
[12] See Appendix A for screenshots of the instructions.
[13] Because we randomly assigned participants into their respective conditions and exclusions were not evenly distributed, sample sizes between conditions, while not identical, are comparable in size.
[14] This is a 10-item questionnaire aimed to capture utilitarian reasoning with the ability to discriminate between individual differences in permissive attitudes towards instrumental harm and impartial concern for the greater good.

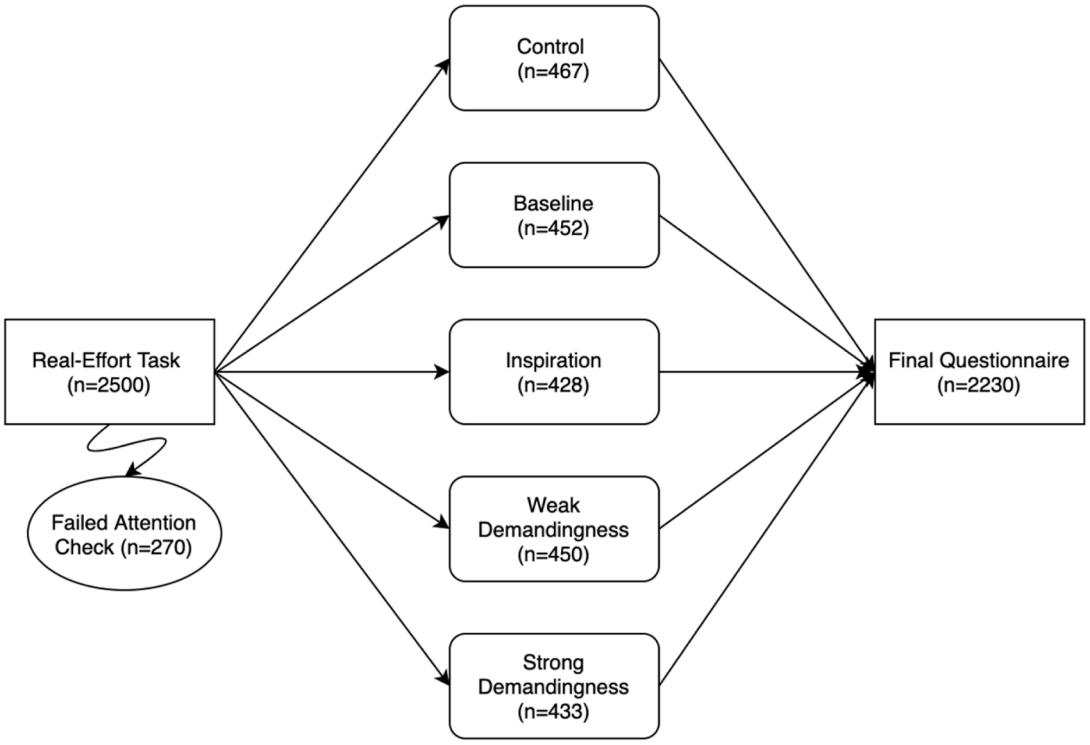

Fig. 3. Experimental Procedure.

Carpenter's (2021) altruism categorization survey, and a few more questions about GiveDirectly and charitable giving more generally. Demographic information for each participant was provided by a demographic presurvey that all Prolific participants undertook before completing any activities on Prolific. Participants earned an average of £1.62 and in total donated £843.15 to GiveDirectly.

2.3. Hypotheses

We investigated two main hypotheses for primary analysis:

Null Hypothesis I. Increasing levels of moral demandingness has no effect on the frequency of giving.
Null Hypothesis II. Increasing levels of moral demandingness has no effect on the total amount of giving.

Given uncertainty about the direction of the effect that moral demandingness may have on donation behavior and because we did not have reason a priori to expect it to increase or decrease giving behavior, the null hypotheses are two-sided, as are our tests.

2.4. Results[15]

2.4.1. Average donation behavior

The main outcome of interest relating to the two null hypotheses is donation behavior, i.e., how many participants made a donation (frequency), and how much of their endowment they donated (amount). Table 1 reports the frequency of donations and the mean amount donated in each condition. We find that there is a significant increase in donation frequency (22.4%) in the Moral Argument treatment compared to the Control (60.0% vs 49.0%, χ^2 test, $p<0.001$). Comparing the Moral Argument treatment to each of the three moral demandingness treatments, we find no statistically significant difference in the frequency of donations (Inspiration vs Moral Argument: 63.1% vs 60.0%, χ^2 test, $p=0.317$; Weak Demandingness vs Moral Argument: 60.0% vs 60.0%, χ^2 test, $p=0.945$; Strong Demandingness vs Moral Argument: 62.6% vs 60.0%, χ^2 test, $p=0.395$).

Table 1
Descriptive Statistics.

	Obs.	*Mean Donation*	*SD*	*Donation*
Control	467	5.42	7.43	49.0%
Moral Argument	452	8.22	8.66	60.0%
Inspiration	428	8.56	8.56	63.1%
Weak Demandingness	450	8.17	8.17	60.0%
Strong Demandingness	433	7.86	7.86	62.6%

Notes: Observations per group, mean donation, standard deviation, and share of participants donating at least 1 ECU.

The mean donation amount in the Control treatment is 5.42 ECUs, which amounts to 27.1% of participants' endowments. This amount is similar to the percentage of the endowment donated found in similar experimental contexts (Eckel and Grossman, 1996; Small and Loewenstein, 2003; van Rijn et al., 2017). We find that participants donate 51.7% more to GiveDirectly in the Moral Argument compared to the Control treatment (8.22 vs. 5.42, two-tailed t-test, $p<0.001$). While donations in our three moral demandingness treatments are also significantly greater than the Control,[16] we find no significant differences between these treatments and the Moral Argument treatment (Inspiration vs Moral Argument: 8.56 vs 8.22, t-test, $p=0.562$; Weak Demand vs Moral Argument: 8.17 vs 8.22, t-test, $p=0.868$; Strong Demand vs Moral Argument: 7.86 vs 8.22, t-test, $p=0.516$).

Table 2 reports regressions of the effect of morally demanding statements on donation behavior compared to the Moral Argument. We fail to find significant effects regarding the frequency of donations in

[15] We present the results from self-reported data on obligation, Altruistic type, and the OUS scale in Appendix C.

[16] Compared to the Control all treatments are significantly greater with $p<0.001$ (two-tailed t-tests).

Table 2
Effect of Moral Demandingness on Donation Behavior.

VARIABLES	(1) Freq.	(2) Amount	(3) Amount>0	(4) Amount>0 (tobit)
Inspiration	0.0329	0.341	-0.176	-0.232
	(0.0329)	(0.577)	(0.605)	(1.156)
Weak Demandingness	-0.00225	-0.0963	-0.109	-0.172
	(0.0326)	(0.570)	(0.607)	(1.160)
Strong Demandingness	0.0279	-0.368	-1.196*	-2.429*
	(0.0328)	(0.575)	(0.605)	(1.141)
Constant	0.600***	8.233***	13.72***	18.39***
	(0.0230)	(0.402)	(0.428)	(0.843)
Observations	1,750	1,750	1,075	1,075
R-squared	0.001	0.001	0.005	

Note: The Constant refers to the Moral Argument. Standard errors in parentheses *** p<0.001, ** p<0.01, * p<0.05. Regression (1) uses a LPM (no difference in results with a probit regression). Regressions (2) and (3) use OLS and regression (4) uses a Tobit model with an upper limit of 20.

regression 1 (LPM model), or the total amount donated in regression 2 (OLS). However, only considering participants who made a donation, we find that participants in the strong demandingness treatment donate significantly less ECUs (1.196) than in the Moral Argument (regression 3). Further, using a tobit model censored at the most that can be donated (20), we find that for participants that did donate, those in the strong demandingness treatment gave 2.429 less ECUs.

These results suggest that moral arguments explaining why one should donate to charity increases the frequency and amount of donations compared to the Control condition. On the other hand, manipulating the level of demandingness in addition to the moral argument does not affect the frequency of donations. While on average, it also does not affect the total amount donated, we do find that conditional on participants donating, they give less in the Strong Demandingness treatment.

2.4.2. Gender differences

We preregistered gender as a variable for secondary analyses. We hypothesized that a gender difference may exist, as females have higher empathic concern than males (Van Rijn et al., 2017) and a higher sense of moral obligation (Einolf, 2011).[17] Fig. 4 plots the mean amount donated by men and women in each treatment. Firstly, we find that compared to the Control, both women and men give significantly more in the Moral Argument treatment (Women: 8.73 vs. 5.58, two-tailed t-test, p<0.001; Men: 7.21 vs. 5.05, two-tailed t-test, p<0.001). Secondly, we fail to find a difference between Moral Argument and Inspiration and Weak Demandingness for men (Inspiration vs Moral Argument: 7.40 vs 7.21, two-tailed t-test, p=0.85; Weak Demand vs Moral Argument: 7.49 vs 7.21, two-tailed t-test, p=0.77) or women (Inspiration vs Moral Argument: 9.13 vs 8.73, two-tailed t-test, p=0.59; Weak demand vs Moral Argument: 8.68 vs 8.73, two-tailed t-test, p=0.95).

However, when comparing the Strong Demandingness and the Moral Argument treatment, we fail to find a significant change in giving (Strong Demandingness vs Moral Argument: 9.46 vs 8.73, two-tailed t-test, p=0.300), though men give significantly less, at around 31% (Strong Demandingness vs Moral Argument: 4.97 vs 7.21, two-tailed t-test, p=0.02).[18]

This pattern also appears in Fig. 5 which reports frequency of giving. We find that more women donate to charity in the Strong Demandingness treatment compared to the Moral Argument treatment (72.4% vs. 63.1%, χ^2 test, p=0.02) while we do not find a significant effect for men (43.9% vs. 53.2%, χ^2 test, p=0.11).[19]

2.5. Discussion

To summarize the findings from the first experiment, we observe that moral arguments increase donations, though varying demandingness on top of this moral argument does not affect donation behavior. However, we do find that participants in the Strong Moral Demandingness treatment, conditional on donating, give less. As a result, we fail to reject Null Hypothesis I as well as Null Hypothesis II, as we do not find a statistically significant change in donor behavior within the moral demandingness treatments in the aggregate.

Note that we do observe substantial heterogeneity between men and women in the responses to strongly morally demanding appeals. In the strong demandingness treatment, we find that men donate significantly less, and women are more likely to donate compared to the Moral Argument. However, since this was a secondary hypothesis, replication is needed to establish this result credibly.

3. Experiment 2

Given the above findings from Experiment 1, we decided to conduct a follow-up experiment. We had four central goals with this second experiment. First, because we did not pre-register the comparison of Control to Moral Argument in Experiment 1, we wanted to test the robustness of this finding. Second, we also aimed to replicate the results from Experiment 1 with respect to the gender effect discussed in 2.4.2, which was also not pre-registered as a primary analysis. As such, we pre-registered this gender-specific hypothesis in our second experiment. Third, we were worried about a potential experimenter demand effect in our design of Experiment 1 (Zizzo, 2010; De Quidt et al., 2019). Specifically, our concern was that participant behavior may be impacted by us—the experimenters—giving the moral demands to donate to charity. In De Quidt et al., (2018) they study experimenter demand effects, and find that their own "strong demand" treatment (where they tell participants "You will do us a favor if…") significantly changed participant behavior. As a result, we decided to decouple the experimenter from the moral demands in the conceptual replication.

We took this opportunity of running a replication study to also improve other aspects of the design. This is why we decided to use a third-party to administer the moral argument and demandingness treatments. We argue that this different context mitigates, at least to a certain extent, possible experimenter demand effects that may arise from us—the experimenters—being the one to provide these arguments to participants in the previous experiment,[20] thus strengthening this study's design. In this follow-up, the third-party non-profit Giving What We Can (GWWC) presents the treatments to participants.[21] As a result, the follow up experiment should be viewed as a conceptual replication of experiment 1. Fourth, we also aimed to replicate the central question of moral demandingness in this new context with a unique focus on strong moral demandingness, as we reasoned that if there was no effect across all versions of moral demandingness, it would be most interesting

[17] Since this effect does not replicate in the second experiment, we do not discuss the gender difference literature in more detail here.

[18] Men also donate significantly less in Strong Demandingness treatment compared to Inspiration (4.97 vs 7.49, two-tailed t-test, p<0.010) and Weak Demandingness treatments (4.97 vs 7.40, two-tailed t-test, p=0.006).

[19] However, less men decide to give to charity in Strong Demandingness treatment compared to Inspiration (43.9% vs. 55.5%, χ^2 test, p=0.050) and significantly less than Weak Demandingness (43.9% vs. 57.1%, χ^2 test, p=0.022).

[20] See footnote 8 for discussion on experimenter demand effects.

[21] GWWC aims to establish a community of effective givers, inspiring donations to the world's most effective organizations. They launched in 2009 and people can sign the GWWC pledge to give 10% of their income to the most effective charities in the world.

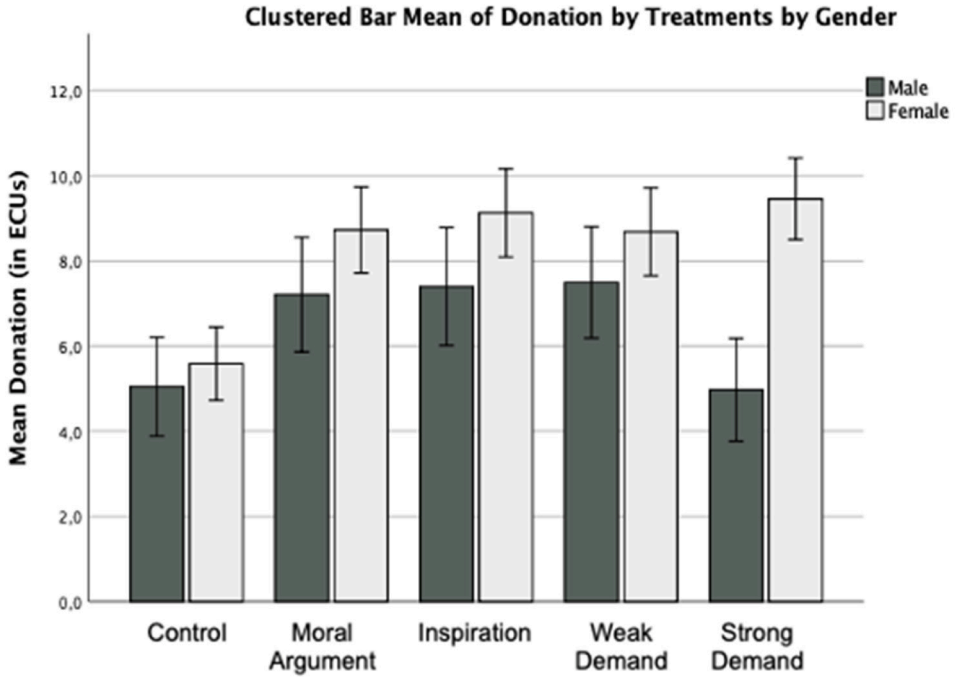

Fig. 4. Mean Donation Amount by Treatment By Gender.
Notes: Mean donations in ECUs by treatment by gender, error bars represent 95% CI.

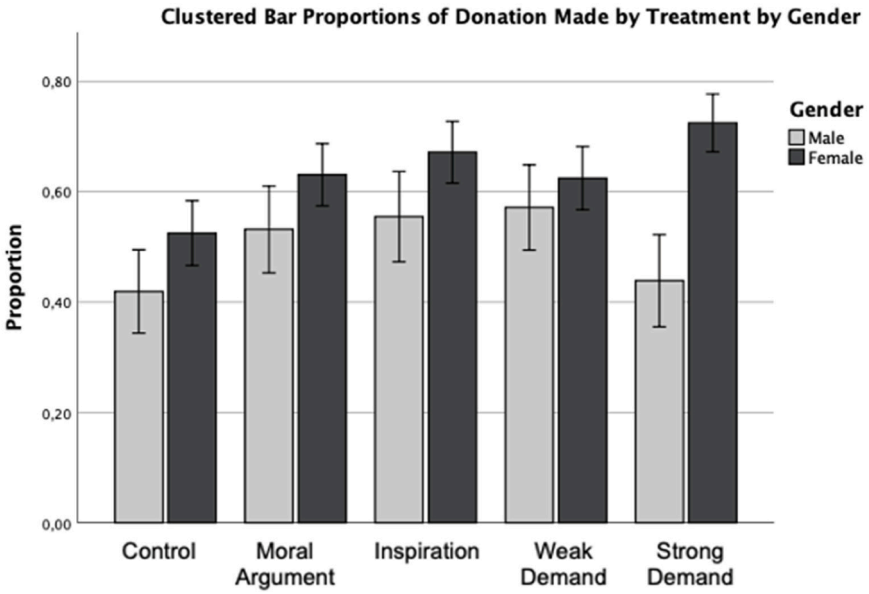

Fig. 5. Proportion of Donations by Treatment By Gender.
Notes: *Frequency of donations is in % by treatment by gender.*

to replicate the results in the context of strong moral demandingness with the highest statistical power within our budgets.

Given these aims, we preregistered the following null hypotheses on OSF.[22] Our null hypotheses are as follows:

Null Hypothesis III. There is no difference in donation behavior between the Control and the Moral Argument treatments.

Null Hypothesis IV. There is no difference in donation behavior between the Moral Argument and the Strong Demandingness treatment.

Null Hypothesis V. There is no difference in donation behavior between men and women within and between the treatments.

3.1. Participants

In the same manner as Experiment 1, we recruited our sample via Prolific. In total, n=1200 participants completed the study. Using G*power, (Faul et al., 2007) our power analysis indicated a required

[22] https://osf.io/hra8w/?view_only=83dfc9832bea421884e29056aa5feb6a.

sample of 400 participants in each treatment, which gave us .8 power to detect an effect as small as .2 (Cohen's d) at a standard .05 alpha error probability. In the experiment, 10.3% of participants failed the attention check and their observations were excluded from all analyses.

3.2. Experimental design

Compared to Experiment 1, the experimental procedure differed on two accounts. First, we only randomly assigned participants into one of three treatments to directly focus on our research goals with Experiment 1: Control, Moral Argument, and Strong Demandingness. Second, instead of presenting the texts to participants within Qualtrics, we provided screenshots from a third-party non-profit (Giving What We Can) with the same texts (see Fig. 6).[23]

Participants were also provided with links to the website displaying these same texts on Givingwhatwecan.org to further clarify that the moral argument and moral demands themselves were not coming from us, the experimenters. To ensure this was the case, we explicitly stated that the screenshots were from GWWC's website, and they could check by clicking the link themselves, thereby verifying our claim. Importantly, the moral arguments were indeed endorsed by GWWC, making our statement truthful.[24] We also elicited personal norms, added measures of guilt proneness, agreeableness, feelings of manipulation, as well as further control variables such as subjective financial well-being, religious affiliation, and religious participation to further understand donor behavior and control for potential confounds. We report these in Appendix E. Participants earned an average of £1.67 and in total donated £394.3 to GiveDirectly.

3.3. Results[25]

Table 3 reports the frequency and mean amount donated. We find the same pattern as in Experiment 1. When comparing the Control to the Moral Argument, we find that people are significantly more likely to donate in the Moral Argument (56.6% vs 44.3%, χ^2 test, $p<0.001$) and donate more on average (7.49 vs 5.24, t-test, $p<0.001$). As a result, we can reject Null Hypothesis III. This remains statistically significant after adjusting for multiple hypothesis testing.[26]

Next, we test Null Hypothesis IV. Once again, we find the same pattern as in Experiment 1. We fail to detect a difference between the Moral Argument and the Strong Demandingness treatments in the frequency of donations (56.6% vs 55.2%, χ^2 test, $p=0.669$) or the average amount donated (7.49 vs 7.00, t-test, $p<0.406$).

Table 4 reports regressions, comparing the effects of the Strong Demandingness statement on donation behavior compared to the Moral Argument. Our results differ from Experiment 1. We fail to find significant differences in any of the regressions. These include the frequency of donations in regression 1 (LPM model), the amount donated in regression 2 (OLS), the amount donated conditional on those who donated in regression 3 (OLS), and the amount donated conditional on those who donated in regression 4 (tobit model with upper limit of 20).

Lastly, we test Null Hypothesis V to see if there is a difference in donation behavior between men and women. Table 5 below reports the donation frequency and amount for men and women in each treatment. When comparing the Moral Argument to the Strong Demandingness treatment, we fail to find a difference in the frequency of donations for men (54.3% vs 53.3%, χ^2 test, $p=0.839$) or women (58.9% vs 56.6%, χ^2 test, $p=0.641$). Furthermore, when we compare these two treatments for the total amount donated, we once again find fail to find a significant difference for men (7.04 vs 6.94, t-test, $p=0.901$) or women (8.01 vs 7.06, t-test, $p=0.294$). These results suggest that unlike Experiment 1, here we fail to find a difference in donation behavior for men and women. As a result, we cannot reject Null Hypothesis V. Tables 6 and 7

In order to provide evidence in favor of a null for results from both of our studies, which is something that regression analyses in a standard null-hypothesis framework technically cannot provide, we report a set of equivalence tests of the regression coefficients based on the approach by Alter & Counsell (2021), which equate to two one-sided tests (TOST) against two equivalence bounds that allow one to draw the conclusion that the estimate is either null or negligibly small, i.e. within these bounds. Below we report this for regression coefficients from both Experiment 1 and Experiment 2, first with regard to probability of donating and then with regard to amount donated. In order to conclude a null effect, both tests against the lower and upper bound have to be significant. We focus on the estimate of our Strong Demandingness dummy as this was our main variable of interest throughout.

Our results indicate that we have evidence in favor of a null effect of the Strong Demandingness treatment in both Study 1 and Study 2 at the equivalence bound level of .1 (in unstandardized coefficients), while we already have evidence for a negligibly small effect at .075 for Study 2. This suggests that in Study 1, the effect of the Strong Demandingness treatment is either null or smaller than a 10% increase or decrease in the frequency of donation compared to the Moral Argument treatment (and 7.5% for Study 2). For the donation amount, we find a null effect or an effect of negligibly small size at the 1.5 level. This suggests that for both studies, the effect of the Strong Demandingness is either null or smaller than a 1.5 ECU increase or decrease in donations. Overall, we take these results to provide evidence in favor of a null effect at this level.[27]

3.4. Discussion

The results of the second experiment indicate that the (now preregistered) results pattern of moral arguments significantly improving upon the control holds, as we also observe such an effect in the aggregate in Experiment 2. Further, we fail to replicate our gender effect from Experiment 1, leading us to conclude that there is no gender effect in the context of morally demanding charitable solicitation. Interestingly, we cannot determine whether the gender effect is not observed in Experiment 2 because it is simply a false positive in Experiment 1, or due to there being experimenter demand effects (EDE) in Experiment 1. While our beliefs lean towards the former, there is an argument to be made for EDE. Across both experiments, donation behavior for men and women are quite similar in the Control Treatment and Moral Argument Treatment. However, there is a divergence in donation behavior for males and females in the strong demandingness treatment respectively. Males donate more than in experiment 1 and females donate less, resulting in their donation behavior on the extensive and intensive margins being very similar in the Strong Demand treatment for Experiment 2 (extensive margin: $p=0.356$; intensive margin $p=0.129$). This analysis is exploratory, given it is difficult to make comparisons across experiments with different population samples (in terms of time). However, given

[23] In the Control treatment, we used a screenshot (and link) of the UK government website, where the original Control text was taken from.

[24] Note that the provision of our materials at Givingwhatwecan.org involve no deception, as all statements are endorsed by the organization and its executive director.

[25] We report the results of personal norms, guilt proneness, agreeableness, feelings of manipulation, as well as regressions with demographic controls in Appendix E.

[26] We provide robustness checks for multiple hypotheses of our pairwise tests based on List et al. (2019). See Appendix F, Table 1, for all our estimations of outcomes. Our results remain significant after adjusting for multiple hypotheses at the 5% level according to List et al. (2019) Theorem 3.1.

[27] We also do not find evidence that demographic controls and personality trait measures meaningfully explain choices with respect to the difference between the Moral Argument and the Control or between the Moral Argument and the Moral Demandingness. For detailed analyses, please see the Online Appendix.

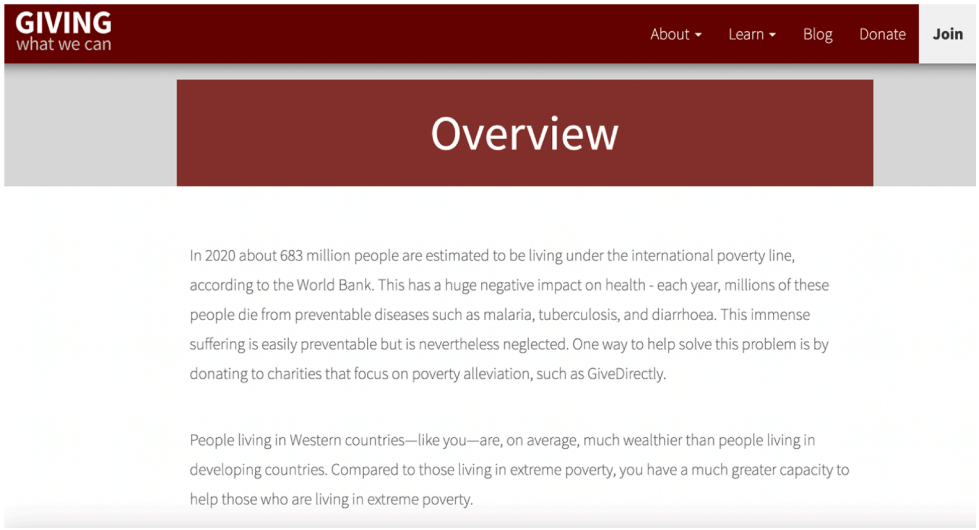

Fig. 6. Moral Demandingness Treatment on GWWC.

Table 3
Donation Behavior.

	Obs.	*Mean Donation*	*SD*	*Donation*
Control	386	5.24	7.47	44%
Moral Argument	417	7.49	8.45	57%
Strong Demandingness	391	7.00	8.26	55%

Notes: Observations per group, mean donation, standard deviation, and share of participants donating at least 1 ECU.

Table 4
Effects of Strong Demandingness on Donation Behavior.

VARIABLES	(1) Freq	(2) amount	(3) amount>0	(4) amount>0 tobit
Strong Demand	-0.0135	-0.489	-0.561	-0.311
	(0.0350)	(0.588)	(0.670)	(0.377)
Constant	0.566***	7.492***	13.24***	5.288***
	(0.0243)	(0.409)	(0.463)	(0.675)
Observations	808	808	452	452
R-squared	0.000	0.001	0.002	

Note: The Constant refers to the Moral Argument. Standard errors in parentheses *** p<0.001, ** p<0.01, * p<0.05. Regression (1) uses a LPM (no difference in results with a probit regression). Regressions (2) and (3) use OLS and regression (4) uses a Tobit model with an upper limit of 20.

Table 5
Donation Behavior by Gender.

	Obs.	*Mean Donation*	*SD*	*Donation*
Men				
Control	172	4.47	7.05	38%
Moral Argument	208	7.04	8.32	54%
Strong Demandingness	210	6.94	8.24	53%
Women				
Control	213	5.89	7.75	50%
Moral Argument	207	8.01	8.58	59%
Strong Demandingness	175	7.06	8.41	57%

Notes: Observations, mean donation, standard deviation, and share of participants donating at least 1 ECU per gender and per group.

how participants change their behavior in the strong demand treatment of De Quidt et al., (2018), it is interesting to observe the differing behavior of males and females when the EDE are mitigated from Experiment 1 to Experiment 2.

Table 6
TOST for Frequency of Donation.

	-.05	.05	-.075	.075	-.1	.1
Strong Demandingness (Exper. 1)	2.38**	.67	3.14***	1.44	3.90***	2.20*
Strong Demandingness (Exper. 2)	1.04	1.81*	1.76*	2.53**	2.47***	3.24***

Notes: All t-statistics for TOST procedures on a variety of lower and upper equivalence bounds (in standardized coefficients) of treatment effects from Models (1)-(5). *** p<0.001, ** p<0.01, * p<0.05.

Table 7
TOST for Amount of Donation.

	-.5	.5	-.1	.1	-.1.5	1.5
Strong Demandingness (Exper. 1)	.02	1.68*	.87	2.53***	1.72*	3.38***
Strong Demandingness (Exper.2)	.23	1.51	1.10	2.38**	1.87*	3.25***

Notes: All t-statistics for TOST procedures on a variety of lower and upper equivalence bounds (in standardized coefficients) of treatment effects from Models (1)-(5). *** p<0.001, ** p<0.01, * p<0.05.

Above and beyond the pre-registered nature of these replications, Experiment 2 also improved the design over the previous iteration, as the treatment texts were presented by a third party. As we also replicated our main result, namely that moral demandingness does not have an additive effect above and beyond moral arguments, and because of the design improvements, we argue that one should have increased confidence in the results of Experiment 2.

4. General discussion

In this paper, we investigate how moral arguments and variations in the level of moral demandingness affect donation behavior. Our experimental design allows us to distinguish the impact that different levels of moral demandingness have on donations.

First, we find that moral arguments in general increase both the frequency of donations and the average amount donated. We replicate this finding in our follow-up experiment. This is in line with previous literature on moral nudges that finds that moral interventions of various kinds, such as teaching moral arguments (Schwitzgebel, Cokelet and

Singer 2020), or presenting participants with statements of moral theories (Dal Bo and Dal Bo, 2014) affect behavior. Our Moral Argument treatment focusses on the participant's relative affluence. Previous studies that simply focus on people's relative rank in the global income distribution provide mixed results: Fehr et al, (2021) find that correcting people about their estimations of their respective global income ranking does not lead to an increase support for global redistribution. However, Nair (2018) finds that people guessing their relative rank and then providing them with information on their true rank can make the information salient, leading to an increase in giving to global charities. Our results suggest that the standard moral argument that informs participants of facts relating to global poverty and points to the outsized impact that those from wealthier countries can have on this poverty is effective at increasing donations. This is similar to the saliency effect that Nair (2018) finds and supports the hypothesis that moral arguments focusing on relative affluence can increase donations to global charities.

Second, we fail to find differences in charitable giving behavior between the treatments that add and vary the levels of moral demandingness. We also replicate this null result in our follow-up and provide exploratory equivalence tests in favor of the null. In short, this result does not support the worries of philosophers who are concerned about a broad backfire effect stemming from strongly demanding claims if we were to communicate these moral obligations (Kagan, 1989; Unger, 1996; De Lazari-Radek and Singer, 2010). Returning to the initial question raised by Sachs (2019), whether we "should demand the demanding", our results indicate that above and beyond making a moral argument, increasing moral demandingness fails to make a difference. Our response to the initial question is that one should not worry too much about the potential backfire effects of morally demanding communication.

Our secondary analyses in the first experiment found behavioral differences between men and women's response to high levels of moral demandingness, which may indicate that one would want to consider differently demanding messages for different populations. However, we fail to replicate this gender difference in our follow-up study. As such, we conclude that there are no gender differences in responses to morally demanding charitable solicitations. While we arrive at this conclusion, an open question remains whether this was due to the gender result being a false positive in Experiment 1, or because of males and females responding differently to experimenter demand effects in Experiment 1.

In simple terms, a well-supported conclusion that we can draw from our findings is that when it comes to demanding the morally demanding, one may want to use moral arguments first and foremost. This is because we have found robust effects of moral arguments across both of our experiments. We make no specific recommendation as to the use of morally demanding language as we did not find any effects. On the other hand, if someone a priori expected a backfire effect from strong morally demanding statements, our results may be seen as evidence in favor of using more demanding statements on the margin than one would otherwise have, but not because there is empirical evidence that it would increase donations, rather because there is no evidence that it has a backfire effect. As this is the case, honestly communicating the moral demandingness may thus be an acceptable approach on authenticity grounds, allowing organizations to honestly communicate the level of moral demandingness they see as central to their claims.

A different way to look at our results is to see our Moral Argument condition as providing participants with context for their donation decision, which provides information as to the need and allows them to make an informed decision as to their ability to help. The moral demandingness treatments, though, may make the demander themselves seem morally superior to the potential donor, which may explain the results presented in this paper. We do note that it is also possible that the effects of our morally demanding statements were watered down by the moral argument itself. People may respond differently to morally demanding statements when they are decoupled from arguments.

Overall, our findings may also have broader potential applicability beyond the context of charitable giving. This is because the general structure of a moral obligation that can be communicated in varying levels of demandingness is present in many domains in the social and political realm, among them climate change activism, the social justice movement, vegetarianism advocacy, or health-related behavior messaging in a pandemic. In all of those, the worry of a backfire effect is present and finding that, on average, increasing levels of moral demandingness does not impact behavior is worth pointing out and may lead to an adaption in communication strategy.[28] We are excited by the prospect of future empirical research on the impact of moral demandingness in these other contexts. We hope that our results add some insight into how to approach these challenges of communicative strategy in the philanthropic sector.

Declaration of Competing Interest

Both authors received grant funding from the Forethought Foundation and the Centre for Effective Altruism. Further, both authors hold fellowships also funded by the Forethought Foundation and the Centre for Effective Altruism. The charity used in this research (GiveDirectly) is recommended as a high-impact non-profit organization by Effective Altruism Funds, which is administered by the Centre for Effective Altruism.

Data availability

We have included our raw data and our analysis (STATA do files) in the submission.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2023.101988.

References

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

Alter, U., & Counsell, A. (2021). Equivalence Testing for Multiple Regression. Available at: https://psyarxiv.com/ugc9e/.

Altmann, S., Falk, A., Heidhues, P., Jayaraman, R., & Teirlinck, M. (2019). Defaults and Donations: Evidence from a Field Experiment. *Review of Economics and Statistics, 101*(5), 808–826.

Bekkers, R. (2005). It's Not All in the Ask. Effects and Effectiveness of Recruitment Strategies Used by Nonprofits in the Netherlands. In *Paper presented at the 34rd Annual ARNOVA Conference*.

Bekkers, R., & Wiepking, P. (2011). A Literature Review of Empirical Studies of Philanthropy: Eight Mechanisms that Drive Charitable Giving. *Nonprofit and voluntary sector quarterly, 40*(5), 924–973.

Berkey, B. (2020). Effectiveness and Demandingness. *Utilitas*, 1–14.

Blumenthal, M., Christian, C., Slemrod, J., & Smith, M. G. (2001). Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota. *National Tax Journal, 54*(1), 125–138.

Böhm, R., Gürerk, Ö., & Lauer, T. (2020). Nudging Climate Change Mitigation: A Laboratory Experiment with Inter-Generational Public Goods. *Games, 11*(4), 1–20.

Bos, B., Drupp, M.A., Meya, J., Quaas, M., 2020. "Moral Suasion and the Private Provision of Public Goods: Evidence from the COVID-19 Pandemic." Available at SSRN 3611579.

Bryant, W. K., Slaughter, H. J., Kang, H., & Tax, A. (2003). Participating in Philanthropic Activities: Donating Money and Time. *Journal of Consumer Policy, 26*(1), 43–73.

Bursztyn, L., Fiorin, S., Gottlieb, D., & Kanz, M. (2019). Moral Incentives in Credit Card Debt Repayment: Evidence from a Field Experiment. *Journal of Political Economy, 127*(4), 1641–1683.

Capraro, V., Jagfeld, G., Klein, R., Mul, M., & van de Pol, I. (2019). Increasing Altruistic and Cooperative Behaviour with Simple Moral Nudges. *Scientific Reports, 9*(1), 1–11.

[28] However, it is important to point out that there are significant heterogeneities between these different contexts, such that a generalisation from our result to other contexts like climate-related behaviour may not be possible without further empirical study.

Carpenter, J. (2021). The Shape of Warm Glow: Field Experimental Evidence from a Fundraiser. *Journal of Economic Behavior & Organization, 191*, 555–574.
Dal Bó, E., & Dal Bó, P. (2014). Do the Right Thing:" The Effects of Moral Suasion on Cooperation. *Journal of Public Economics, 117*, 28–38.
De Lazari-Radek, K., & Singer, P. (2010). Secrecy in Consequentialism: A Defence of Esoteric Morality. *Ratio, 23*(1), 34–58.
Eckel, C. C., & Grossman, P. J. (1996). The Relative Price of Fairness: Gender Differences in a Punishment Game. *Journal of Economic Behavior & Organization, 30*(2), 143–158.
Eckel, C. C., & Grossman, P. J. (2003). Rebate Versus Matching: Does how we Subsidize Charitable Contributions Matter? *Journal of Public Economics, 87*(3-4), 681–701.
Edwards, J. T., & List, J. A. (2014). Toward an Understanding of why Suggestions Work in Charitable Fundraising: Theory and Evidence from a Natural Field Experiment. *Journal of Public Economics, 114*, 1–13.
Einolf, C. J. (2011). Gender Differences in the Correlates of Volunteering and Charitable Giving. *Nonprofit and Voluntary Sector Quarterly, 40*(6), 1092–1112.
Ekström, M. (2021). The (Un)compromise Effect: How Suggested Alternatives can Promote Active Choice. *Journal of Behavioral and Experimental Economics, 90*, Article 101639.
Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A Flexible Statistical Power. *Analysis Program for the Social, Behavioral, and Biomedical Sciences. Behavior Research Methods, 39*(2), 175–191.
De Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review, 108*(11), 3266–3302.
De Quidt, J., Vesterlund, L., & Wilson, A. J. (2019). *Experimenter demand effects. In Handbook of research methods and applications in experimental economics*. Edward Elgar Publishing.
Fehr, D., Mollerstrom, J., & Perez-Truglia, R. (2022). Your place in the world: Relative income and global inequality. *American Economic Journal: Economic Policy, 14*(4), 232–268.
Gandullia, L. (2019). The Price Elasticity of Warm-Glow Giving. *Economics Letters, 182*, 30–32.
Gandullia, L., Lezzi, E., & Parciasepe, P. (2020). Replication with MTurk of the experimental design by Gangadharan, Grossman, Jones, and Leister (2018): charitable giving across donor types. *Journal of Economic Psychology*, Article 102268.
Gandullia, L., & Lezzi, E. (2018). The Price Elasticity of Charitable Giving: New Experimental Evidence. *Economics Letters, 173*, 88–91.
Giamattei, M., Yahosseini, K. S., Gächter, S., & Molleman, L. (2020). LIONESS Lab: A Free Web-Based Platform for Conducting Interactive Experiments Online. *Journal of the Economic Science Association, 6*(1), 95–111.
Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than do Subject Pool Participants. *Behavior Research Methods, 48*(1), 400–407.
Huck, S., & Rasul, I. (2010). Transactions Costs in Charitable Giving: Evidence from Two Field Experiments. *The BE Journal of Economic Analysis & Policy, 10*(1).
Ito, K., Ida, T., & Tanaka, M. (2018). Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand. *American Economic Journal: Economic Policy, 10*(1), 240–267.
Kagan, S. (1989). *The Limits of Morality*. New York: Oxford University Press.
Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond Sacrificial Harm: A Two-Dimensional Model of Utilitarian Psychology. *Psychological Review, 125*(2), 131–164.
Karlan, D., & List, J. A. (2007). Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment. *American Economic Review, 97*(5), 1774–1793.
List, J. A., Murphy, J. J., Price, M. K., & James, A. G. (2019). Do Appeals to Donor Benefits Raise More Money than Appeals to Recipient Benefits? Evidence from a Natural Field Experiment with Pick.Click.Give. (No. w26559). *National Bureau of Economic Research*.
List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics, 22*(4), 773–793.
MacAskill, W. (2015). *Doing Good Better: Effective Altruism and a Radical New Way to make a Difference*. Guardian Faber Publishing.
Meer, J. (2011). Brother, Can You Spare a Dime? Peer Pressure in Charitable Solicitation. *Journal of Public Economics, 95*(7-8), 926–941.
Meer, J., & Rosen, H. S. (2012). Does Generosity Beget Generosity? Alumni Giving and Undergraduate Financial Aid. *Economics of Education Review, 31*(6), 890–907.
Nair, G. (2018). Misperceptions of Relative Affluence and Support for International Redistribution. *The Journal of Politics, 80*(3), 815–830.
Palan, S., & Schitter, C. (2018). Prolific. ac—A Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27.
Pummer, T. (2016). Whether and Where to Give. *Philosophy & Public Affairs, 44*(1), 77–95.
Reiley, D., & Samek, A. (2019). Round Giving: A Field Experiment On Suggested Donation Amounts In Public-Television Fundraising. *Economic Inquiry, 57*(2), 876–889.
Sachs, B. (2019). Demanding the Demanding. In T. Pummer, & H. Greaves (Eds.), *Effective Altruism: Philosophical Issues* (pp. 137–149). Oxford: Oxford University Press.
Schwitzgebel, E., Cokelet, B., & Singer, P. (2020). Do Ethics Classes Influence Student Behavior? Case Study: Teaching the Ethics of Eating Meat. *Cognition, 203*, Article 104397.
Shang, J., & Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal, 119*(540), 1422–1439.
Singer, P. (1972). Famine, Affluence, and Morality. *Philosophy & Public Affairs*, 229–243.
Torgler, B. (2004). Moral Suasion: An Alternative Tax Policy Strategy? Evidence from a Controlled Field Experiment in Switzerland. *Economics of Governance, 5*(3), 235–253.
Unger, P. K. (1996). *Living High and Letting Die: Our Illusion of Innocence*. Oxford University Press.
Van Rijn, J., Barham, B., & Sundaram-Stukel, R. (2017). An Experimental Approach to Comparing Similarity-and Guilt-Based Charitable Appeals. *Journal of Behavioral and Experimental Economics, 68*, 25–40.
Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*(1), 75–98.