Cost-(in)effective public good provision: an experimental exploration

Nathan W. Chan[1] · Stephen Knowles[2] · Ronald Peeters[2] ·
Leonard Wolk[3]

Abstract

This paper investigates the determinants of cost-(in)effective giving to public goods. We conduct a pre-registered experiment to elucidate how factors at the institutional and individual levels shape individual contributions and the cost-effectiveness of those contributions in a novel public good game. In particular, we examine the role of consequential uncertainty over the value of public good contributions (institutional level) as well as individual characteristics like risk and ambiguity attitudes, giving type, and demographics (individual level). We find cost-ineffective contributions in all institutions, but total contribution levels and the degree of cost-ineffectiveness are similar across institutions. Meanwhile, cost-effectiveness varies by giving type—which is a novel result that is consistent with hypotheses we generate from theory—but other individual characteristics have little influence on the cost-effectiveness of contributions. Our work has important positive and normative implications for charitable giving and public good provision in the real world, and it is particularly germane to emerging online crowdfunding and patronage platforms that confront users with a multitude of competing opportunities for giving.

Keywords Risk · Ambiguity · Cost-effectiveness · Public goods · Charity

✉ Ronald Peeters
ronald.peeters@otago.ac.nz

Nathan W. Chan
nchan@umass.edu

Stephen Knowles
stephen.knowles@otago.ac.nz

Leonard Wolk
l.wolk@vu.nl

[1] Resource Economics, University of Massachusetts Amherst, Stockbridge Hall, 80 Campus Center Way, Amherst, MA 01003, USA

[2] Department of Economics, University of Otago, PO Box 56, Dunedin 9054, New Zealand

[3] Department of Finance, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

⚫ Springer

1 Introduction

Individuals engage with myriad public goods in their daily lives, ranging from environmental quality to human services to public spaces and public art, and they may contribute to these various causes through monetary donations, in-kind gifts, and volunteerism. This diverse landscape presents private citizens with a rich choice set for public good contributions, but it also creates challenges for coordination and cost-effective allocation of resources. Given vast heterogeneity in public goods and technologies for augmenting public goods, there is latitude for misallocation of resources across causes (Chan & Wolk, 2020). These issues have become even more pronounced in recent times, as online crowdfunding and patronage platforms offer a growing menu of public good possibilities to a rapidly expanding base of contributors.

In this paper, we investigate the determinants of cost-(in)effective contributions to public goods, exploring factors at both the institutional and individual levels. At the institutional level, we examine whether patterns of giving are influenced by consequential uncertainty over the value of public good contributions. At the individual level, we consider risk and ambiguity attitudes, giving type as elicited through a charitable giving task, and common demographic covariates. To understand these relationships, we model behavior in a public good game that accounts for heterogeneity in giving types and social preferences. We derive from this model testable predictions, and we proceed to test these hypotheses in a pre-registered online experiment.

Our experimental design builds on that of Chan and Wolk (2020), which features a set of four simultaneous public good contribution decisions with different marginal per capita returns (MPCRs). Importantly, this multiple public good environment makes possible cost-ineffective contributions, as subjects may contribute at low MPCRs without exhausting all contribution possibilities at high MPCRs. We extend this framework to examine how individual and institutional factors affect the cost-effectiveness of contributions. In particular, we construct two treatments in which the MPCR of each public good is a random variable with known bounds. Importantly, bounds are non-overlapping across the four public goods, allowing us to construct a clean and novel measure of cost-ineffectiveness to test our hypotheses. In one treatment, the distributions for these random variables are known (Risky treatment), while in another treatment, the distributions are unknown (Ambiguity treatment); we compare these treatments to each other and to a control treatment in which MPCRs are certain (Certain treatment). We also include an array of additional tasks to elicit individual characteristics, allowing us to investigate the relationship between cost-ineffectiveness and demographics, risk and ambiguity attitudes, risk literacy, giving types, and attentiveness.

We construct a measure of cost-ineffectiveness that conditions on the subject's total contribution. In this way, the measure focuses on how well the subject allocates the contribution across disparate public goods; the measure should *not* be directly affected by the subject's cooperativeness or generosity and is thus well-suited for our subsequent analysis. We conduct a series of parametric and non-parametric tests to

compare contribution behavior across treatments and across individuals. Cost-ineffective giving is present in all three treatments, indicating that subjects are not maximizing the impact and payoffs generated by their chosen level of contributions. Interestingly, total contributions and the degree of cost-ineffectiveness are nearly identical across treatments, which suggests that the information environment does not affect these margins of decision-making. Turning to individual characteristics, we find little evidence that risk or ambiguity attitudes affect cost-ineffectiveness. However, there are some differences across giving types. In particular, we find that individuals who are at least partially motivated by warm glow (i.e., pure warm-glow givers and impure altruists) contribute less cost-effectively than pure altruists and non-donors. This finding accords with predictions from theory, although we are the first, to our knowledge, to document this empirically. Our findings are robust for within- and between-subjects analyses.

Our experimental results provide important insights for the broader world. Individuals have always had many avenues for augmenting public goods, e.g., when facing multiple charitable causes. These choice sets continue to expand with modern crowdfunding and patronage platforms, making room for inefficient allocation of resources across public goods. Inefficiency may arise from individual or institutional factors, and understanding the influence of these different factors is crucial to improving public good provision. In many cases, there may be uncertainty about the productivity of different public good investments. For example, how well do investments in civic crowdfunding projects (e.g., urban greenspaces, public art installations, etc.) enhance social cohesion? How well do individual actions (e.g., wearing face masks, minimizing social contact) help blunt the spread of infectious diseases? There are also important questions about how public good contributions differ along dimensions of individual heterogeneity. What types of individuals are most likely to contribute in cost-(in)effective ways? Overall, our work provides novel and timely insights into both positive and normative aspects of public good provision.

Our work advances several distinct lines of inquiry in the economics of public good provision and charitable giving. In focusing on efficiency, our work ties in with discussions spurred by the Effective Altruism movement. Effective Altruists urge donors to give to charities that generate the greatest benefit per dollar donated. This idea has been a topic of substantial interest among philosophers and ethicists (MacAskill, 2016; Singer, 2015), but it also has natural intersections with efficiency and cost-effectiveness concerns long articulated by economists (Karlan & Wood, 2017; List, 2011). These principles have also gained traction outside of academia, e.g., in the form of organizations like Charity Navigator and GiveWell, which seek to quantify the social impact of dollars donated across charities and causes.

Yet, in spite of clear economic and ethical rationales for effective altruism, many individuals still appear unresponsive or inattentive to the effectiveness of their charitable efforts. Evidence from laboratory and field experiments shows that most donors do not increase their donations when provided with information about the effectiveness of the charity (Clark et al., 2018; Karlan & Wood, 2017). Metzger & Günther (2019) show that many participants in laboratory experiments are unwilling to purchase, even for a minimal fee, information on the efficiency of the charity they

have been asked to donate to.[1] Instead, these participants show greater interest in purchasing information about the people that the charity will help, suggesting that efficiency is not a top priority. Along similar lines, Berman et al. (2018) report that, for most participants, emotional attachment to a charitable cause is more important than the effectiveness of the charity, which may explain why few people behave as effective altruists. Genç et al. (2020) show via a choice experiment that most donors place significantly greater weight on where a donation is spent (preferring it to be spent closer to home), while assigning less importance to the effectiveness of the donation or the needs of the recipient. In short, there is robust evidence that many donors show little interest in the efficiency of the charities to which they are donating.

We explore how this (in)attention to cost-effectiveness may relate to institutional and individual factors. Our investigation into individual characteristics contributes to a broader literature on how contribution behaviors differ across motivations for giving. An influential body of theory within economics separates motives for giving into two main types: pure altruists and warm-glow givers (Andreoni, 1989, 1990; Ribar & Wilhelm, 2002; Warr, 1982; Yildirim, 2014). Pure altruists derive utility from the total amount of the charitable good provided. To a pure altruist, her donation and the donations of others are perfect substitutes, leading to crowding out of charitable donations (Warr, 1982). In contrast, a warm-glow giver derives utility from the act of giving itself, thus removing the scope for crowding out (Andreoni, 1989, 1990). A third category, impure altruists, comprises individuals who earn utility both from own donations and total donations. It is often assumed that warm-glow givers are more prone to inefficient donations than pure altruists (Singer, 2015), a result that follows clearly from theory. However, to our knowledge, there is no direct empirical evidence supporting this claim. For example, Null (2011) separates donors into different types by *assuming* those who donate inefficiently must be warm-glow givers (after ruling out risk aversion). Similarly, Karlan and Wood (2017) report that some participants increase donations when presented with information on the effectiveness of donations, but many participants do not. They posit that the latter group may include warm-glow givers, but their experimental design does not allow them to ascertain this. Against this backdrop, our work provides an important contribution by demonstrating a direct link between giving type and cost-effective public good provision.

Individuals have a wide array of opportunities to contribute to public goods and charitable causes, so an implicit concern raised in all of the aforementioned studies is that donors may misallocate their resources toward less worthy causes. A number of recent studies have tackled this issue of multiple public goods directly. Much of the work in this realm focuses on coordination of donors across public goods. Corazzini et al. (2015) show that coordination problems can be eliminated by making one public good focal by offering better payoffs. Earlier work by Cherry and Dickinson (2008) similarly finds that subjects successfully coordinate on the option with highest

---

[1] Gangadharan and Nemes (2009) report a similar finding in the context of a public good game: when there is an unknown probability that the private account and/or public account will not be paid out, most participants choose not to pay a small fee to find out what this unknown probability is.

social returns when faced with multiple public goods, even when the level of social returns is endogenous to aggregate contributions. Interestingly, when comparing a setting with multiple homogeneous public goods to an equivalent setting with a single public good, they report greater contributions in the former than the latter, suggesting the importance of framing. Bernasconi et al. (2009) also investigate this "unpacking" effect and similarly find improvements in contribution levels in the unpacked case. Similar to the above papers, Blackwell and McKee (2003) construct an environment with global and local public goods. They find that when the two goods provide equal societal returns, subjects donate more to the local public good; however, when social returns are higher for the global cause, subjects give more to the global public good—in spite of it generating smaller private returns.[2]

We build on this tradition by implementing an experiment with multiple public goods, but we highlight a different aspect of the problem. Extending the design of Chan and Wolk (2020), we eliminate the scope for coordination problems, thus allowing sharper focus on individual allocations as the locus of inefficiency. In this way, our experiment more directly addresses issues raised by effective altruists, who express concerns about individuals' willingness to allocate funds to inferior causes. Although our work bears similarities to Chan and Wolk (2020), it is distinct. Whereas Chan and Wolk (2020) provide a proof-of-concept for this experimental design and shed light on framing effects induced by the choice sets, we provide deeper insight into policy relevant determinants of cost-ineffectiveness. We show how the propensity for cost-(in)effective donations may be influenced by individual characteristics and the broader information environment.

Our findings on the provision of inferior public goods also shed light on the industrial organization of charities. In particular, how do less efficient and less impactful charities survive in the market? We find that warm-glow givers and impure altruists often spread their contributions across causes, including inferior ones that might not otherwise survive under effective altruism. This pattern of behavior can uphold otherwise unproductive charities, with accompanying implications for efficiency of charity markets and social welfare.

Incorporating uncertainty has been a topic of interest in the public good games literature. However, while a number of papers study risk in public good games [see, e.g., Dickinson (1998), Gangadharan and Nemes (2009), Freundt and Lange (2021) and Théroude and Zylbersztejn (2020)], few study ambiguity. Levati and Morone (2013) and Björk et al. (2016) are among the few to examine both risk and ambiguity, and they do not find significant differences in contribution behavior in situations involving risky, ambiguous, or deterministic MPCRs for single public good settings. According to the latter, there is also no interaction between strategic uncertainty and natural uncertainty. They find that cooperative attitudes and beliefs about group members' contributions are unaffected by natural uncertainty. Even so, focus has remained on settings with single public goods, which does not allow for studying

---

[2] While there is some literature studying giving to multiple *charities* [e.g. Meer (2017)], public good games are a different context as "the dictator game [that is typically used in the context of charitable giving] does not capture how the incentive to free ride affects the interaction between potential donors" (Vesterlund, 2015, p. 94).

cost-effectiveness. Our work thus provides novel insights into the interplay between cost-effectiveness and the overarching information environment.

The remainder of the paper proceeds as follows: Section 2 provides a general model, lays out hypotheses for our public good game, and describes details of our experimental implementation. We discuss results in Sect. 3, while Sect. 4 concludes.

2 Experimental design and hypotheses

2.1 Design

We begin by laying out a general model of our game environment. Let there be n players and m public goods with prices normalized to unity. Following Chan and Wolk (2020), each player i has budget $w_j^i > 0$ to allocate to public good j: $x_j^i \in [0, w_j^i]$. The marginal per capita return (MPCR) of public good j is denoted γ_j. That is, player i's contribution to public good j, x_j^i, produces a benefit of $\gamma_j \cdot x_j^i$ to all players $k = 1, \ldots, n$. For each public good $j = 1, \ldots, m$, we assume $\gamma_j \in (0, 1)$, such that not contributing to any of the public goods is the only rationalizable strategy for a selfish payoff maximizing player.

Departing from Chan and Wolk (2020), let the MPCRs be subject to risk: $\gamma_j \sim F_j$. Let $\widetilde{\gamma}_j = \int_{\gamma_j} \gamma_j dF_j$ and public goods be ordered such that $\widetilde{\gamma}_1 < \cdots < \widetilde{\gamma}_m$. Finally, let the support of γ_j be $(\underline{\gamma}_j, \overline{\gamma}_j)$.[3]

Defining utility over total payoffs, the expected utility of player i is given by

$$\int_{\gamma_1} \cdots \int_{\gamma_m} u^i \left(\sum_{j=1}^m \left[\left(w_j^i - x_j^i \right) + \gamma_j \cdot \left(x_j^i + X_j^{-i} \right) \right] \right) dF_m \cdots dF_1,$$

where $X_j^{-i} = \sum_{k \neq i} x_j^k$. Rewriting this expression as

$$\int_{\gamma_1} \cdots \int_{\gamma_m} u^i \left(\sum_{j=1}^m \left[w_j^i - \left(1 - \gamma_j \right) \cdot x_j^i + \gamma_j \cdot X_j^{-i} \right] \right) dF_m \cdots dF_1,$$

we see that both the effective cost $1 - \gamma_j$ and the effective benefit γ_j are subject to risk.

If a player is risk-neutral, we derive the same condition as Chan and Wolk (2020) for the cost-effective allocation of resources. That is, if $x_j^i > 0$, then cost-effectiveness requires that $x_\ell^i = w_\ell^i$ for all $\ell > j$; otherwise, this player can increase her ex ante expected utility by shifting resources from x_j^i to x_ℓ^i. Clearly, this holds in an example where $u^i(\pi) = \pi$. In this case, the utility expression from above can be simplified to:

---

[3] To ensure the social dilemma inherent in public goods games, we constrain the support so that each F_j is within $(\frac{1}{n}, 1)$.

$$\int_{\gamma_1} \cdots \int_{\gamma_m} \sum_{j=1}^{m} \left[w_j^i - \left(1 - \gamma_j\right) \cdot x_j^i + \gamma_j \cdot X_j^{-i} \right] dF_m \cdots dF_1$$

$$= \sum_{j=1}^{m} \left[w_j^i - \left(1 - \widetilde{\gamma}_j\right) \cdot x_j^i + \widetilde{\gamma}_j \cdot X_j^{-i} \right].$$

Alternatively, if player i is risk-averse (e.g., $u^i(\pi) = \pi^\alpha$ with $\alpha \in (0, 1)$), this is not obvious. The reason is that payoffs may depend on player i's beliefs about the contributions of others (i.e., X_j^{-i} for all $j = 1, \ldots, m$). For instance, if i expects all others to contribute fully to public good m ($x_m^k = w_m^k$) and nothing to the other public goods ($x_j^k = 0$ for $j < m$), player i may decide to contribute a positive amount to public good $m - 1$ instead of m to mitigate risk. This possibility arises when γ_m and γ_{m-1} have overlapping supports, so that γ_{m-1} may be realized at a higher value than γ_m.

We conduct our experiment with $n = 3$ players and $m = 4$ public goods. For each public good, players have an endowment of $w_j^i = 10$ points available that they can contribute to public good j. We implement three treatments with differing marginal per capita returns (MPCR), γ_j, for each public good:

Treatment certain. The values of the four MPCRs are *certain*:

$$\gamma_1 = 0.475 \quad \gamma_2 = 0.625 \quad \gamma_3 = 0.775 \quad \gamma_4 = 0.925.$$

Treatment risky. The values of the four MPCRs are subject to *risk*:

$$\gamma_1 \sim \mathrm{Un}(0.40, 0.55) \quad \gamma_2 \sim \mathrm{Un}(0.55, 0.70) \quad \gamma_3 \sim \mathrm{Un}(0.70, 0.85) \quad \gamma_4 \sim \mathrm{Un}(0.85, 1.00).$$

Treatment ambiguity. The values of the four MPCRs are subject to *ambiguity*:

$$\gamma_1 \in (0.40, 0.55) \quad \gamma_2 \in (0.55, 0.70) \quad \gamma_3 \in (0.70, 0.85) \quad \gamma_4 \in (0.85, 1.00).$$

We specify disjoint supports in the latter two treatments (with $\underline{\gamma}_\ell > \overline{\gamma}_j$ for all $\ell > j$), which ensures that there is a clear ranking in the four MPCRs ($\gamma_4 > \gamma_3 > \gamma_2 > \gamma_1$) in all three treatments. This ensures that, regardless of their risk attitude, cost-effectiveness of contribution behavior is well-defined in all three treatments: Individuals who contribute cost-effectively would not contribute to a public good with lower MPCR unless they exploited their full endowment in all public goods with higher MPCR. That is, $x_j^i > 0$ only if $x_\ell^i = w_\ell^i$ for all $\ell > j$.

2.2 Hypotheses

In our pre-registration, we log the following study objective: "Our primary outcome is whether uncertainty of consequences in public goods affect cost-effectiveness of individual contributions." As our secondary outcome, we ask: "How does cost-effectiveness of contributions relate to individual attitudes toward risk and ambiguity, and giving type?"

Our primary null hypothesis is that players' contributions are cost-effective in all treatments (regardless of their beliefs about the other players' choices).

Hypothesis 1 Contributions will be cost-effective in all treatments.

For the secondary objective, we focus on two types of individual characteristics: risk and ambiguity attitudes and giving type. Because we have designed the treatments so that MPCR supports are disjoint, risk and ambiguity attitude should not yield differences in the cost-effectiveness of contributions.

Hypothesis 2 In all treatments, cost-effectiveness of contributions do not correlate with risk or ambiguity attitude.

We now consider the role of giving type. We consider four giving types as defined by Gangadharan et al. (2018) and Gandullia et al. (2020): non-donors, pure altruists, warm-glow givers, and impure altruists. The latter three types may contribute to public goods, but for different reasons. Above, we assumed that players' utilities are defined over total payoffs to allow for more parsimonious exposition of risk types. Here, we consider richer preference structures to elucidate differences across giving types.

Pure altruists may contribute because they derive utility from the aggregate level of public good provision, e.g.,

$$u^i \left(\underbrace{\sum_{j=1}^{m} \left[w_j^i - x_j^i \right]}_{\text{numeraire}}, \underbrace{\sum_{j=1}^{m} \gamma_j \left[x_j^i + X_j^{-i} \right]}_{\text{public good}} \right). \tag{1}$$

Such behavior would also be consistent with a preference for efficiency, and there is evidence from prior public good experiments that (some) subjects behave in such a way (Goeree et al., 2002).

However, there is also evidence from public good experiments of warm-glow givers who instead derive utility from their own act of giving (Andreoni, 1993). These individuals may value contributions to each public good separately, e.g.,

$$u^i \left(\underbrace{\sum_{j=1}^{m} \left[w_j^i - x_j^i \right]}_{\text{numeraire}}, \underbrace{x_1^i, \ldots, x_m^i}_{\text{contributions to each m}} \right), \tag{2}$$

or they may derive utility from the total amount they have contributed across all m public goods, e.g.,

$$u^i \left(\underbrace{\sum_{j=1}^{m} \left[w_j^i - x_j^i \right]}_{\text{numeraire}}, \quad \underbrace{\sum_{j=1}^{m} x_j^i}_{\text{total contributed}} \right). \tag{3}$$

Impure altruists combine both warm-glow and altruistic motives. Importantly, these differences in preference structures across giving types have implications for cost-(in)effective giving. A pure altruist has no reason to contribute cost-ineffectively to the public good, as doing so will reduce the total amount of the public good provided. On the other hand, a warm-glow giver may contribute cost-ineffectively. If there are diminishing marginal warm glow benefits for each specific public good in (2), then a warm-glow giver may spread their contributions across public goods in a cost-ineffective manner. Indeed, this is an underlying assumption of Null (2011). However, even a warm-glow giver who values total contributions over all public goods could act similarly. In this case, contributions to $m - 1$ and m are perfect substitutes in (3), and the individual will be indifferent between contributing to one cause or the other, which can beget a cost-ineffective allocation across causes.

As a result, we expect more cost-effective contributions among pure altruists and more cost-ineffective contributions among warm-glow givers, with impure altruists falling between these two poles. We summarize these insights in the following hypothesis:

Hypothesis 3 Pure altruists will contribute in a cost-effective manner. Impure altruists will contribute less cost-effectively than pure altruists, and warm-glow givers will contribute even less cost-effectively than impure altruists.

Because our experimental design directly manipulates uncertainty across treatments, we are best positioned to make causal statements on the first primary hypothesis. We do not experimentally manipulate risk and ambiguity attitudes or individual giving types, so our conclusions on the last two secondary hypotheses will be based on correlational evidence.

2.3 Procedures

On Tuesday 25 August 2020 and Monday 24 April 2023, we invited up to, respectively, 216 and 108 potential participants (18–65 years, fluent in English) via *Prolific* to participate in a "decision-making experiment".[4] In total, respectively, 201

---

[4] Prolific is increasingly being used by researchers for online experiments [see, e.g., Chan and Wolk (2020), Granulo et al. (2019) and Hafner et al. (2019), among others]. The quality of online samples has been confirmed through multiple direct studies (Arechar et al., 2018; Casler et al., 2013; Crump et al., 2013). Regarding Prolific in particular, Peer et al. (2017) engaged with researchers to identify several priority areas for data quality: comprehension, attention, honesty, and reliability. Comparing MTurk, Prolific, and online panels (like Qualtrics Panels) on these dimensions, the authors find that Prolific had superior performance, especially in terms of comprehension, attention, and dishonesty. Gupta et al. (2021) find Prolific is slightly noisier than a traditional lab experiment, but given the cost per subject, Prolific dominates from an inferential power perspective. They find MTurk to be much noisier than Prolific.

and 107 participants both accepted the consent form and completed all tasks.[5] All these participants visited all three treatments and completed four public good tasks in each treatment, but the order in which they visited the three treatments was subject to individual randomization. After completing the experimental treatments, subjects completed a post-experiment questionnaire that comprised a sequence of short individual tasks in which we elicited gender, age, giving type (Gangadharan et al., 2018; Gandullia et al., 2020), risk attitude (Eckel & Grossman, 2002, 2008), ambiguity attitude (Baillon et al., 2018), attention level (Frederick, 2005; Sirota and Juanchich, 2018), and risk literacy (Cokely et al., 2012). We offer brief descriptions of these tasks in the next section and provide full details in Appendix A.

Groups were formed as soon as a triple of participants completed the full suite of tasks.[6] Each treatment was completed as a one-shot game.[7] The three players within a group were paid according to the same randomly drawn treatment (although players may have visited those treatments in a different sequence) and for each public good according to the same (randomly drawn) MPCR, and received feedback on their final earnings at the end of the experiment.[8] All this was common knowledge to the participants. On average, participants earned £4.08 in variable payment, in addition to the participation fee (which was £4.00 in the first session and £5.00 in the second session) for on average 23 min and 27 s of their time. Subjects were informed of their

---

[5] Our original study with 201 participants would allow us to identify treatment effect sizes of about 0.5 at the 5% level and with 80% power in a between-subjects analysis (computed using a two-sample t-test) for our primary hypothesis, and a within-subject analysis would allow us to identify an effect size of about 0.2 (computed using a one-sample t-test). This justified our earlier choice to aim for at least 60 observations per treatment. In response to a reviewer's request for conceptual replication, we decided to expand our sample by 50%, which would allow us to identify an effect size of about 0.33 in a between-subjects analysis and of about 0.16 in a within-subject analysis. Given that, unlike in the original study, no order effects are found, the expansion of the data set allows us to move from a between-subject analysis to a within-subject analysis, thereby decreasing the minimum detectable effect size from 0.5 to 0.16. In only two cases are our observed effect sizes smaller than what our power analysis allows us to identify in the within-subjects treatment comparisons: (1) Total contributions between certain and ambiguous; (2) Relative cost-ineffectiveness (RCI) between risk and ambiguous. Additionally, for full transparency, the between-subject analysis is presented in the appendix where all observed effect sizes for the treatment comparisons are larger than the thresholds computed using the power analysis.

[6] In theory, such a procedure could lead to endogenous selection. However, we do not believe this to be a viable concern in the Prolific environment. Upon launch, the experiment becomes available to all eligible users on Prolific. This is a fluid subject pool that is far larger than our final enrollment, and users may log on or notice the newly posted experiment at various times for idiosyncratic reasons; users also do not know the extent or capacity of our experimental enrollment beyond the two other subjects they have been partnered with. Thus, it is very unlikely that users who arrive at the beginning are systematically different than those who arrive minutes later.

[7] We focus on a one-shot setting to remove the element of coordination among players that arises in repeated VCM games. Coordination would complicate predictions and introduce potentially confounding variation, particularly in a setting with cost-ineffective contribution mechanisms, ambiguity, and risk. Thus, the use of a one-shot game helps focus attention on the results of interest. One-shot games are common in the public goods literature, including notable works by Goeree et al. (2002), Walker and Halloran (2004), and Drouvelis and Grosskopf (2016).

[8] To compute the MPCRs for the Ambiguity treatment, we draw for each public good j two values a_j and b_j from a uniform distribution, using the clocktime at which the first participant entered the group as the seed. Next, for each public good j, the MPCR γ_j is the random draw from the $\beta(a_j, b_j)$ distribution, scaled to the respective MPCR interval.

earnings for all tasks at the end of the experiment; importantly, this means they were unaware of their earnings from the public goods game when they completed the giving-type elicitation.[9]

Prior to data collection, ethical approval was obtained from Vrije University School of Business and Economics Research Ethics Review Board (reference code SBE6/9/2020kwk350) and University of Otago's Human Ethics Committee (reference code D20/183). Moreover, this study was pre-registered in advance on 18 August 2020 in the AEA RCT Registry under the unique identifying number "AEARCTR-0006304". The experiment was programmed using *oTree* (Chen et al., 2016). Screenshots are available in Appendix C.

One advantage of the online setting is that we can track reading times for each page in the study. We find substantive engagement throughout the study. On average, subjects spent 5 min and 46 s on the instructions for our main experimental task and an additional 3 min and 24 s on the tasks themselves. There were also no noticeable drops in engagement over the course of the study. The residence times on each page of our post-experimental questionnaire suggest reflective engagement with each, with longer durations for more complicated questions. For example, subjects spent the most time on our ambiguity attitude questions (4 min and 22 s, on average) and risk literacy questions (2 min and 4 s, on average), the latter of which was the final task in the study. In total, subjects spent over 23 min and 27 s, on average, on the study. Together, these data speak to the attentiveness of our sample. Furthermore, in line with best practices for online experiments, we required subjects to complete several control questions as comprehension checks before proceeding to the main experimental tasks. The comprehension checks for the public goods games were set up as four simultaneous questions presented on a single screen: Three true/false questions and one calculation question (see Appendix C for screenshots of the questions). Participants had one chance to get the questions right. If a participant answered a question incorrectly, we showed an explanation of the correct answer after which they needed to revise their answer to the correct one before proceeding. The percentages of participants who correctly answered the comprehension checks are 76.6% (Q1, true/false), 79.6% (Q2, true/false), 28.6% (Q3, true/false), and 47.7% (Q4, open answer, correct answer is both the mode and median). With the exception of question 3 which was worded in a way that unless a participant paid very close attention there is a chance to answer it incorrectly, participants appear to comprehend the rules of the game well.[10]

---

[9] We acknowledge that participants who gave a lot in the public good game may have depleted their warm glow and as a result have shown less generous behavior in the giving type elicitation task than they would have shown without the public good game, such that a slight misclassification relative to Gangadharan et al. (2018) cannot be excluded.

[10] We do not find a significant difference between cost-ineffectiveness of contributions between participants who answered at least three out of the four questions correctly and those who didn't, nor between participants who answered the fourth question correctly and those who didn't. Although we find significant differences in their total contributions, they do not differ in a way that distorts our cost-ineffectiveness measure. Hence, the results we report are not driven by comprehension at the moment of answering the questions.

3 Results

3.1 Descriptive statistics

Our post-experimental questionnaire included survey questions for basic demographic variables (gender and age) and a series of tasks to elicit each participant's giving type, risk attitude, ambiguity attitude, risk literacy, and attentiveness. We summarize these tasks here briefly and provide additional details on each in Appendix A.

For giving type, we use a two-stage charitable giving task, following the design of Gangadharan et al. (2018) as implemented by Gandullia et al. (2020). Based on the donations made in each stage of this task, we classify each participant as a non-donor, a pure warm-glow donor, a pure altruist, or an impure altruist. To elicit risk attitudes, we use the lottery menu described by Eckel and Grossman (2002, 2008), and we use the method of Baillon et al. (2018) to elicit ambiguity attitudes. We normalize both scales so that they range from 0 to 1, with low values indicating risk/ambiguity-aversion and high values indicating risk/ambiguity loving attitudes. Attentiveness was measured with a cognitive reflection test (CRT) (Frederick, 2005) using the multiple choice format described by Sirota and Juanchich (2018). We use the four questions from the Berlin Numeracy Test of Cokely et al. (2012) to measure risk literacy. Both attentiveness and risk literacy are encoded as the fraction of correct answers given to these survey questions.

Table 1 presents participant characteristics based on the post-experimental questionnaire. We show statistics for each treatment sequence (as described above)

Table 1 Participant characteristics

Seq	N	Gender		Age	Giving type				Risk	amb	crt	r.lit
		Male (%)	Female (%)		None (%)	p.w-g (%)	p.alt (%)	i.alt (%)				
CRA	45	62.2	37.8	26.6	0.0	28.9	4.4	66.7	0.38	0.49	0.48	0.33
CAR	45	48.9	51.1	26.6	8.9	31.1	4.4	55.6	0.31	0.45	0.41	0.38
RCA	48	58.3	41.7	27.9	6.3	37.5	10.4	45.8	0.46	0.52	0.41	0.32
RAC	65	58.5	41.5	27.3	10.8	23.1	10.8	55.4	0.45	0.49	0.49	0.34
ACR	55	50.9	49.1	27.7	10.9	25.5	9.1	54.6	0.42	0.51	0.42	0.29
ARC	50	57.1	42.9	26.9	2.0	28.0	8.0	62.0	0.43	0.48	0.39	0.35
Total	308	56.0	44.0	27.2	6.8	28.6	8.1	56.5	0.41	0.49	0.44	0.33

Treatment sequences use the following abbreviations: Certain (C), Risky (R), and Ambiguity (A)

Giving types are non-donors (none), pure warm-glow givers (p.w-g), pure altruists (p.alt), and impure altruists (i.alt). The columns for risk, amb, crt, and r.lit provide information on risk attitude, ambiguity attitude, attentiveness, and risk literacy, respectively. The scales for risk and amb are normalized to range from 0 to 1, with low values indicating risk/ambiguity-aversion and high values indicating risk/ambiguity-loving attitudes. The crt and r.lit values represent the fraction of correct answers on the Cognitive Reflection Test and the risk literacy survey, respectively

and for the full sample; we use the following abbreviations: Certain (C), Risky (R), and Ambiguity (A). Kruskal–Wallis tests do not indicate significant differences in participants' gender, age, total donations in the giving-type task, risk attitude, ambiguity attitude, attention level, and risk literacy across the orders ($p > .33$).[11]

We now turn to participant behavior in the experiment. Table 2 presents average contribution levels across the four different public goods, by treatment environment and by sequence. Kruskal–Wallis tests do not identify any order effects in any of the three treatments for participants' contributions to individual public goods or in participants' aggregate contributions over public goods ($p > .31$).

In terms of times required for participants to fill out the decision screens, average completion times are quite similar across treatments (63 s for Certain, 68 s for Risk and 73 s for Ambiguity). However, when further splitting each treatment by order, a Kruskal–Wallis test identifies highly significant differences for the three treatments across orders ($p = .0001$). Further inspection reveals that the observed difference arises because participants progressively decreased completion times over the course of the experiment, likely due to the familiarity with the rules of the game that varied only along a single dimension across the three screens. On average, the first task required 116 s to complete, the second task required 51 s to complete, and the third task required 37 s to complete. Kruskal–Wallis tests do not reveal significant differences across orders for the time taken to complete the first, second and third task ($p > .11$).

Overall, these various analyses give us confidence that there are no order effects. Thus, to generate our primary results, we pool the data and conduct a within-subjects analysis. However, for the sake of transparency and robustness, we replicate all major analyses using a between-subjects approach in Appendix B.

3.2 Contribution behavior and cost-effectiveness

We begin by presenting results on basic contribution behavior to foreground subsequent results on cost-(in)effectiveness. For clarity, we refer to the Wilcoxon signed-rank test as the Wilcoxon test and the Mann–Whitney U (or Mann–Whitney–Wilcoxon) rank-sum test as the Mann–Whitney test throughout the paper.

Table 3 shows, for each of the three treatments, the participants' average contributions to each of the four public goods and aggregated over the four public goods. Wilcoxon tests comparing contributions between consecutive public goods reveal that in all three treatments subjects are responsive to the MPCR ($p = .0000$), with average contributions increasing in MPCR.

Importantly, participants' aggregate contributions over the four public goods do not vary significantly across treatments, as can be seen in Fig. 1, which shows the

---

[11] Because we ran the study in two separate sessions, as described above, it is also worth examining whether there were session effects. We do not find session effects for any of the following variables: gender, risk attitude, ambiguity attitude, total donation in the giving-type task, cognitive reflection test, risk literacy, and the total contributions in the three treatments ($p > .19$). The only session effect is the age of the participants ($p = .0002$), an effect that disappears when correcting for the time difference between the two sessions were run ($p = .9758$). We can safely pool data across sessions without making the age adjustment.

Table 2 Summary statistics for contribution levels

Seq	Certain				Risky				Ambiguity			
	PG 1	PG 2	PG 3	PG 4	PG 1	PG 2	PG 3	PG 4	PG 1	PG 2	PG 3	PG 4
CRA	4.51	5.00	6.11	6.69	4.82	5.09	5.80	7.09	4.62	5.33	5.71	6.80
CAR	4.11	4.73	5.73	6.98	3.84	4.78	5.62	7.00	4.02	4.64	5.82	6.98
RCA	4.48	4.92	5.88	6.75	3.63	4.35	5.10	6.63	4.17	4.54	5.40	6.71
RAC	3.52	4.34	5.43	6.65	3.54	4.14	5.37	6.78	3.49	4.28	5.48	6.75
ACR	4.04	4.78	5.96	6.64	3.78	4.71	5.73	6.82	3.98	4.76	5.56	7.11
ARC	3.92	4.34	5.30	6.42	4.04	4.68	5.46	6.42	3.94	4.60	5.54	6.64
Total	4.06	4.66	5.72	6.68	3.91	4.59	5.51	6.78	4.00	4.67	5.57	6.83

Table 3 Contributions across treatments

Treatment	PG 1	PG 2	PG 3	PG 4	Total
Certain	4.0584	4.6623	5.7175	6.6786	21.1169
Risky	3.9091	4.5942	5.5065	6.7825	20.7922
Ambiguity	4.0000	4.6656	5.5747	6.8312	21.0714

Fig. 1 Total contributions
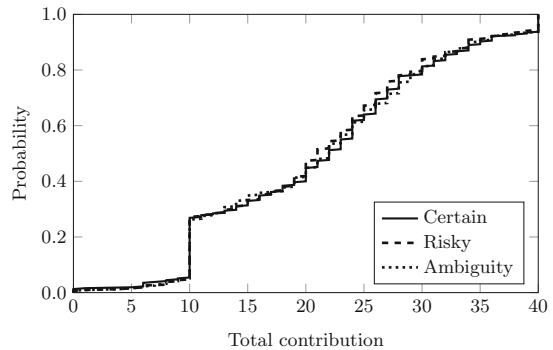
cumulative distributions over these total contributions for each of the treatments. Given the almost perfect overlap of the three curves, it is no surprise that neither a Kruskal–Wallis test for mean equivalence by treatment ($p = .8476$) nor Wilcoxon tests reveal significant differences in total contributions across treatments ($p > .48$). In terms of contributions to individual public goods, Kruskal–Wallis tests do not return any significant differences between treatments ($p > .63$).

We now move to study our main research question: Whether cost-effectiveness differs across treatments. We begin by defining cost-ineffectiveness as

$$\mathrm{CI}(x^i) = \sum_{j=2}^{m} \min\left\{ \sum_{\ell=1}^{j-1} x_{\ell}^i; \sum_{\ell=j}^{m} \left(w_{\ell}^i - x_{\ell}^i \right) \right\},$$

where $w_{\ell}^i = 10$ and $m = 4$ in our experiment. The minimum value for $\mathrm{CI}(x^i)$ is 0 (no ineffectiveness) and the maximum is 40, obtained with the allocation (10, 10, 0, 0).

Notably, the maximal attainable CI varies with the total contribution amount. For example, an individual who contributes only 1 unit could have CI = 3 if she contributes that one unit to PG1. If she contributed 2 units, and both to PG1, should would instead have CI = 6. Because total contributions are endogenous, this property jeopardizes between-subject comparisons. To correct for this, we develop a measure of relative cost-ineffectiveness.

Take an individual who contributes $X^i = \sum_{j=1}^{m} x_j^i$ in total. The maximum cost-ineffectiveness is obtained by this amount being contributed to the least effective public goods. This is the allocation $y^i(X^i)$ defined as

$$y_j^i(X^i) = \max\left\{ \min\left\{ X^i - \sum_{\ell=1}^{j-1} w_{\ell}^i; w_j^i \right\}; 0 \right\}.$$

Now, we can define relative cost-ineffectiveness as

$$\mathrm{RCI}(x^i) = \mathrm{CI}(x^i) / \mathrm{CI}\left(y^i\left(\sum_{j=1}^{m} x_j^i \right) \right).$$

This measure is attractive because it conditions on the total amount contributed, thus allowing for comparison across subjects.

To provide intuition on the process of calculating CI and RCI, we offer an example in Fig. 2. The individual in the left panel of this figure contributes 2 units to PG1, 5 units to PG2, 7 units to PG3 and 9 units to PG4. This individual contributes in total 23 units to the four public goods. If this individual would have been providing these 23 units in the most cost-effective manner, she would have contributed according to the contribution profile in the middle. Starting from the left panel, achieving the cost-effective contribution profile would require moving CI = 7 units, as shown in the figure. The right panel shows the contribution profile if she were to contribute those 23 units in the most cost-*ineffective* manner. To achieve a cost-effective contribution profile from the right panel would require a movement of CI = 37 units. Thus, the relative cost-ineffectiveness of the scheme provided by the individual is RCI = 7/37 ≈ 0.1892.

As noted above, a key advantage of RCI is that the measure is normalized, allowing for comparison across subjects with different total contributions. On the other hand, one challenge with RCI is that it is undefined for individuals who cannot contribute cost-ineffectively, i.e., non-contributors and full contributors.[12] However, practically speaking, this does not pose any problems for our analysis. We find that

---

[12] In our sample, there are four subjects who do not contribute to any public good in at least one of the three treatments and there are twenty subjects who contribute to the full endowment to all public goods in at least one of the three treatments.
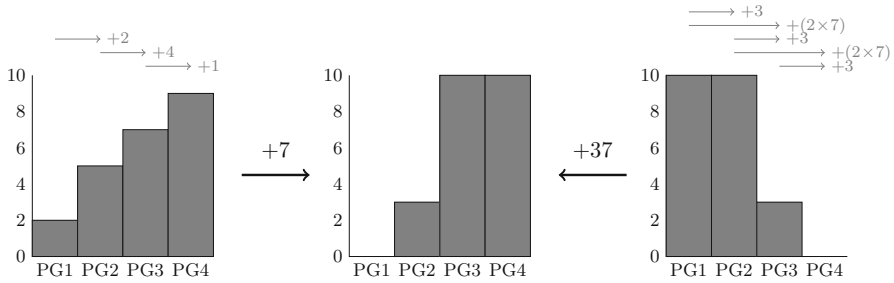
Fig. 2 Example for RCI

total contributions are quite similar across treatments (see distributions in Fig. 1), indicating that total contribution behavior is not a source of significant confounding variation in RCI.

Figure 3 plots the cumulative distributions of RCI for the three treatments. Clearly, cost-ineffective giving is prevalent, indicating that subjects are not maximizing the impact and payoffs generated by their total contributions, which is at odds with Hypothesis 1. However, the prevalence of cost-ineffectiveness is similar across treatments. Given that the three curves in the figure are nearly identical, it is no surprise that neither a Kruskal–Wallis test ($p = .8100$) nor Wilcoxon tests show significant differences across treatments (C vs. R: $p = .1965$; C vs. A: $p = .2823$; R vs. A: $p = .8867$). Thus, the information environment appears to have little bearing on the cost-effectiveness of public good contributions.

3.3 Individual characteristics

We next study how individual characteristics affect the cost-effectiveness of individual contributions which is relevant for Hypotheses 2 and 3. Table 4 presents regression results relating RCI to individual characteristics in each treatment (columns 1–3) and based on the average RCI across treatments (column 4). In our

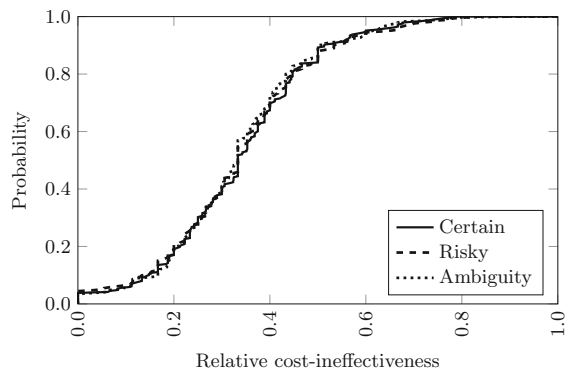Fig. 3 Relative cost-ineffectiveness

Table 4 Relative cost-ineffectiveness

Characteristic	Certain	Risky	Ambiguity	Average
Constant	0.2583***	0.2216***	0.2297***	0.2386***
Gender	0.0354*	0.0575***	0.0420**	0.0415**
Age	0.0017	0.0008	0.0012	0.0012
Giving type				
Pure warm-glow	0.0424	0.0768*	0.0835**	0.0673*
Pure altruist	0.0151	0.0318	0.0571	0.0377
Impure altruist	0.0381	0.0670*	0.0968**	0.0688*
Risk attitude	−0.0185	0.0085	−0.0606**	−0.0234
Ambiguity attitude	0.0341	0.0629	0.0459	0.0479
Attention level (CRT)	−0.0773***	−0.0742***	−0.0497**	−0.0706***
Risk literacy	0.0208	0.0055	−0.0277	−0.0027
Observations	282	287	285	279

*10%; **5%; ***1%. Observations do not include those that do not contribute to any public good and those that contribute fully to every public good; we also exclude subjects for whom other questionnaire data is missing. The reference giving type is the non-donor

discussion of these results, we will evaluate differences at 5% significance, although the table includes markers for significance at the 1% and 10% levels, as well.

With respect to Hypothesis 2, we do not find that ambiguity attitude has any affect on RCI in any of the treatments. Likewise, we find little evidence that risk attitude affects RCI, except in the Ambiguity treatment.

Interestingly, related to Hypothesis 3, we find in all treatments that pure warm-glow givers, pure altruists, and impure altruists are more cost-ineffective compared those who did not donate in the giving-type task (non-donors); however, this effect is only significant in the Ambiguity treatment for donors who are at least in part motivated by warm-glow (pure warm-glow givers and impure altruists). Note that these regression-based results rely upon comparisons against the omitted category (non-donors), which may not be the comparisons of primary interest. Kruskal–Wallis tests find differences in RCI across giving types in both Risky ($p = .0381$) and Ambiguity ($p = .0153$), but not in Certain ($p = .4056$). For Risky and Ambiguity, Dunn's tests indicate a significant difference between impure altruists and pure altruists, between impure altruists and non-donors, and between pure warm-glow givers and non-donors. Figure 4 shows the distribution of the average RCI for each giving type. The average RCI is 0.2525 for non-donors, 0.3369 for warm-glow givers, 0.2905 for pure altruists, and 0.3494 for impure altruists. A Kruskal–Wallis test identifies a significant difference in the average RCI across giving types ($p = .0425$). Subsequent Dunn's tests find the average RCI to be significantly different between warm-glow givers and non-donors ($p = .0219$) and between

Fig. 4 Average relative cost-ineffectiveness and giving type
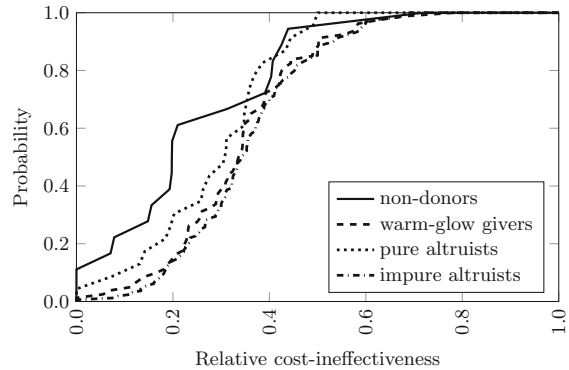
impure altruists and non-donors ($p = .0060$); the difference between impure altruists and pure altruists is only significant at the 10% level ($p = .0513$) and no significant difference is found for the other three pairwise comparisons ($p > .13$). These results suggest that types who are at least partially motivated by warm glow (pure warm-glow givers and impure altruists) are contributing less cost-effectively than those who are not (non-donors and pure altruists). This result can be further corroborated with a Mann–Whitney test comparing the average RCI between these two groups of types ($p = .0083$). We note that these differences in RCI across giving types are not caused by differences in total contributions. Kruskal–Wallis tests give $p = .6131$ on total contributions averaged over the three treatments and $p > .29$ for each treatment separately. Also, average total contributions are not different between the two groups of types (Mann–Whitney: $p = .2057$).

Regarding other individual characteristics, we find some gender differences, with females contributing in a less cost-effective manner compared to males in Risky and Ambiguity. Second, participants with a lower attention level, as measured via the cognitive reflection test, are found to be less cost-effective in all three treatments. Third, risk literacy shows no significant effect.

The between-subjects analysis largely confirms findings of the within-subjects analysis. For Hypothesis 1, we observe a slightly larger dispersion for the RCI between treatments, yet this difference is not significant, in line with the within-subjects analysis. For Hypothesis 2, we do not find significant coefficients for risk or ambiguity attitudes, except in one case: the regression coefficient for ambiguity attitude is positive and significant at the 5% level in Risky. However, if we were to use a more stringent threshold that corrects for multiple hypothesis testing, this effect would be insignificant. With respect to giving types and differences in cost-effectiveness (Hypothesis 3), we again see a similar pattern to that of the within-subjects analysis, but the p-value is larger and now stands at .0773 from a Kruskal–Wallis test. Further detail is available on the between-subjects analysis in Appendix B.

4 Discussion and conclusion

We set out to investigate how cost-effectiveness varies across information environments and by individual characteristics. We find no evidence that the information environment affects overall contributions or the cost-effectiveness of contributions. In terms of individual characteristics, we find that risk and ambiguity attitudes have little bearing on the cost-effectiveness of contributions. Meanwhile, giving type does influence cost-effectiveness, with individuals who are at least partially motivated by warm glow (i.e., pure warm-glow givers and impure altruists) contributing less cost-effectively than those who are not motivated by warm glow (i.e., pure altruists and non-donors).

This latter finding can be rationalized with theory and is consistent with common wisdom concerning warm-glow motives. Indeed, previous literature assumes warm-glow givers are more likely to donate inefficiently than are pure altruists [see, e.g., Null (2011) and Singer (2015)], yet we are not aware of any prior work that directly tests this relationship. Against this backdrop, our results are novel and provide important empirical backing for this commonly held claim. In addition to documenting differences in cost-effectiveness across giving types, we furthermore show that these differences are most pronounced in settings where there may be risk or ambiguity surrounding the value of public good contributions.

More broadly, our experimental work elucidates key factors underlying—or undermining—effective altruism. We shed new light on how the cost-effectiveness of public good provision efforts is influenced by individual characteristics in general and giving types in particular. Our inquiry into environments with multiple public goods is especially germane given expanding options for public good provision through traditional charities and emerging crowdfunding and patronage platforms.

One point about our giving type results merits further discussion. Although we find pure altruists have a lower RCI than warm-glow givers and impure altruists, it may seem surprising that pure altruists display any RCI at all. Theoretically, a pure altruist should consider their contributions and others' contributions as perfect substitutes *and* care about efficiency as they derive utility from the total amount of the public good (or charitable good) provided. Methods for eliciting giving type focus on only one of these two features of pure altruism. Those relying only on perfect substitutability include Crumpler and Grossman (2008), Gangadharan et al. (2018), Ottoni-Wilhelm et al. (2017), and Fielding et al. (2022). Gangadharan et al. (2023) propose a different elicitation method that classifies subjects as pure altruists if their decisions in an experiment show they care about the effectiveness of their donation. We find that some subjects who are classified as pure altruists based on the Gangadharan et al. (2018) elicitation do contribute in a cost-ineffective manner; thus, a subject could be classified as a pure altruist by one of Gangadharan et al.'s classification systems, but not the other.

We can imagine several avenues for future research. First, our theoretical exposition suggests that those with warm-glow motives are more likely to contribute in a cost-ineffective manner. Indeed, we find evidence of this in our experiment, but it is possible that there are other factors that correlate with giving type that drive these between-treatment differences. For example, different giving types may also have different norms or beliefs about how one should allocate resources across different charitable causes. Exploring these possibilities will provide a better understanding of whether giving type, per se, is responsible for our experimental results. A second interesting line of inquiry is to vary whom the individual interacts with in the different public goods. In our experiment, a subject interacts with the same set of group members across all four public goods, but in real-world settings, it is more likely that they will interact with different individuals or networks in each (Bramoullé and Kranton, 2007; Richefort, 2018). Third, there is a need for further inquiry into how giving types are characterized and whether these differ across contexts. In our experiment, we have elicited giving types using a charitable giving task, and we have found that these giving types are correlated with different patterns of behavior in a public good game. However, it is not obvious that giving types will necessarily be consistent across these settings Chan et al. (2023).

Appendix

Short individual tasks: description and summary statistics

Giving type

We use the design of Gangadharan et al. (2018) as implemented by Gandullia et al. (2020). Participants are given 20 points and have to decide how many (if any) of these points, g_1, they want to donate to their preferred charity (out of Oxfam, Red Cross, Save the Children, World Wildlife Fund, and Doctors without Borders), knowing that any amount not donated by them will be donated by us (the

Fig. 5 Giving type
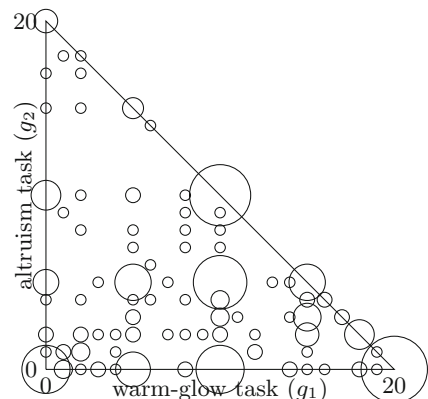
experimenters), such that the charity organization will always receive a total donation of 20 points. After they have made this decision, they are informed that they have $20 - g_1$ points left, and are given the opportunity to donate any amount, g_2, of these to the charity, knowing that this time no further donation will be made by us, and the charity organization will receive in total $20 + g_2$ points.

Figure 5 illustrates the choices of the participants, where the volume of the bubbles are proportional to the number of participants making a particular choice. We use the values g_1 and g_2 to determine a participant's giving type as follows:

g_1	g_2	Giving type
$= 0$	$= 0$	Non-donor or pure selfish
> 0	$= 0$	Pure warm-glow giver
$= 0$	> 0	Pure altruist
> 0	> 0	Impure altruist

The experiments of Gandullia et al. (2020) were conducted via MTurk with 1062 individuals participating. The resulting distribution over the four giving types was 33%, 20%, 8%, and 39% in their study. The distribution resulting from our experiment is 7%, 29%, 8%, and 56%. That is, substantially less non-donors and substantially more impure altruists. In our analysis, we use giving type as a categorical variable.

Risk attitude

We follow the method introduced by Eckel and Grossman (2002, 2008) and ask participants to choose which lottery they want to play out of a menu of lotteries. All lotteries on the menu yield a high payoff, π_H, or a low payoff, π_L, each with a 50–50 chance, but the values of the high and the low payoff differ across lotteries. In order to obtain more variation in the data, participants in our experiment are given the following menu of 11 lotteries (A–K), and associated CRRA intervals:

Lottery		A	B	C	D	E	F	G	H	I	J	K
π_H		10	12	14	16	18	20	22	24	26	28	30
π_L		10	9	8	7	6	5	4	3	2	1	0
CRRA	min	4.91	1.64	1.00	0.72	0.56	0.45	0.37	0.30	0.24	0.16	–
	max	–	4.91	1.64	1.00	0.72	0.56	0.45	0.37	0.30	0.24	0.16

Figure 6 illustrates the choices participants made. In our analysis we take risk as a linear variable and assign the value 0 to the most risk-averse choice (lottery A) and the value 1 to the most risk-loving choice (lottery K).

Fig. 6 Risk attitude
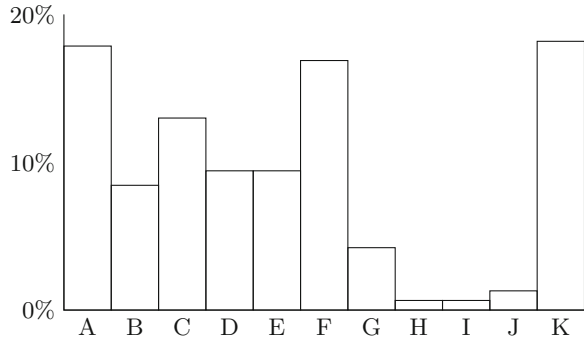
Ambiguity attitude

We use the method of Baillon et al. (2018). An urn is filled with red, green and blue balls.[13] Via a stochastic BDM we elicit reservation prices m_r, m_g and m_b for the option on winning (20 points) by drawing a ball of one particular color, and reservation prices m_{-r}, m_{-g} and m_{-b} for the option on winning (20 points) by drawing a ball of one out of two possible colors. Take $\overline{m}_1 = \frac{1}{3}(m_r + m_g + m_b)$ and $\overline{m}_2 = \frac{1}{3}(m_{-r} + m_{-g} + m_{-b})$. Now, $a = 1 - (\overline{m}_1 + \overline{m}_2) \in [-1, 1]$ measures ambiguity attitude. In order to prevent issues related to curvature of utility functions over points, reservation prices are implemented as probabilities to win the 20 points. In order to keep the experiment within time limits, participants are asked to state the reservation for one one-color option and for one two-color option; the exact colors are randomly drawn at the individual level, but always such that for each participant the two colors in the two-color option are complementary to the color in the one-color option.

The left panel of Fig. 7 shows the participants' choices for m_1 and m_2. The average value of m_1 and m_2 are 45.20 and 57.00. While the average value of m_1 is above the ambiguity-neutral value of 1/3, the average value of m_2 is below the ambiguity-neutral value of 2/3. Although one-third of the participants reported a value of m_2 below that of m_1, overall the values for m_2 are significantly above those of m_1 (Wilcoxon: $p = .0000$). Further, the reported values of m_1 and m_2 correlate significantly (Spearman: $\rho = 0.2186$, $p = .0001$).

The right panel of the figure shows the distribution of $m_1 + m_2$ that we use to evaluate individuals' ambiguity attitude. Values of $m_1 + m_2$ above 100 indicate ambiguity-aversion and those below 100 ambiguity-loving. The average value of $m_1 + m_2 = 102.20$ indicates slight ambiguity-aversion on average. There are 118 participants (38.31%) with $m_1 + m_2 < 96$, 52 (16.88%) with $m_1 + m_2 \in [96, 104]$, and 138 (44.81%) with $m_1 + m_2 > 104$. In our analysis, we encode ambiguity attitude by $1 - \frac{m_1 + m_2}{200}$, such that extreme ambiguity-aversion is at 0, ambiguity-neutral is at 1/2 and extreme ambiguity-loving is at 1.

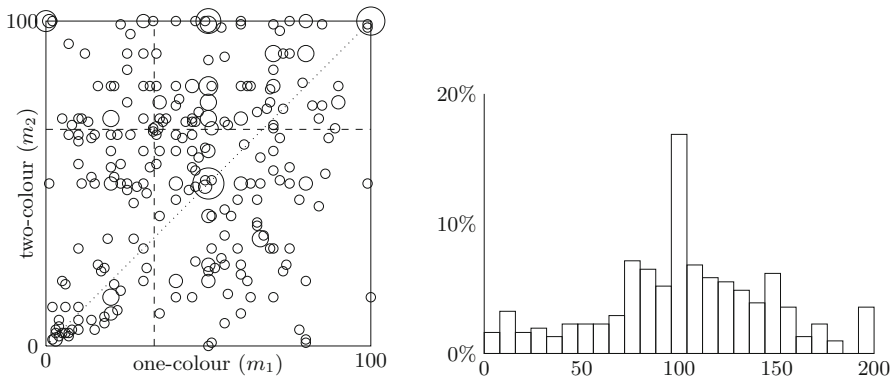[13] We use a method similar to that described in Footnote 8 to design the ambiguous urn.

Fig. 7 Ambiguity attitude. Left: scatter plot of participants' choices (m_1, m_2); right: distribution of $m_1 + m_2$

Attention level and risk literacy

To elicit the participants' attention level, we use the Cognitive Reflection Test of Frederick (2005), where we follow Sirota and Juanchich (2018) in having participants choose from four possible answers—one of the wrong answers being the intuitive one and one being the correct one. Participants are given 90 s to answer the three questions. The percentage of correct answers for the three questions were 38%, 49%, and 51%, while the percentage of intuitive answers for each of the three questions were 60%, 37%, and 39% in Sirota and Juanchich (2018). The respective percentages for our participants, conditional on the questions being answered, are 41%, 51%, and 43% for correct answers and 55%, 33%, and 36% for intuitive answers—not too different from Sirota and Juanchich (2018). The distribution of participants over the number of correctly answered questions is presented in Fig. 8. In our analysis, we incorporate attention level as a linear variable coded as the fraction of correct answers given.

We use the four multiple choice questions from the Berlin Numeracy Test of Cokely et al. (2012) to elicit participants' risk literacy. Participants are given 150 s to answer the four questions. The distribution of participants over the number of correctly answered questions is presented in Fig. 8. The four individual questions were answered correctly, conditional on the questions being answered, in 56%, 43%, 30%, and 15% of the cases, respectively. In our analysis, we incorporate risk literacy as a linear variable encoded via the fraction of correct answers given.

Between-subjects analysis

To complement our primary within-subjects results, we also present between-subjects analysis here. For this analysis, we use only contribution decisions of participants in the first task they visit. This leads to 90 observations for Certain, 113 for Risky and 105 for Ambiguity.

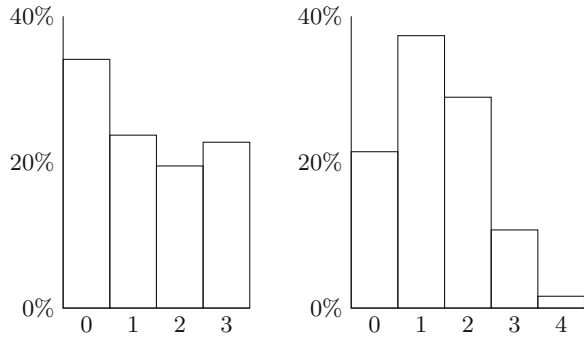Fig. 8 Attention level (left) and risk literacy (right)

Table 5 Contributions across treatments

Treatment	PG 1	PG 2	PG 3	PG 4	Total
Certain	4.3111	4.8667	5.9222	6.8333	21.9333
Risky	3.5752	4.2301	5.2566	6.7168	19.7788
Ambiguity	3.9619	4.6857	5.5524	6.8857	21.0857

Contributions

Average contributions are displayed in Table 5, which is directly comparable to Table 3 for our within-subjects analysis. As before, we find average contributions increasing with MPCR in all treatments ($p = .0000$). Kruskal–Wallis tests do not identify differences between treatments for any of the four individual public goods ($p > .22$) or for total contributions ($p = .2702$). The cumulative distributions over these total contributions for each treatment are shown in Fig. 9, which is the between-subjects analog of Fig. 1 presented in the main text.
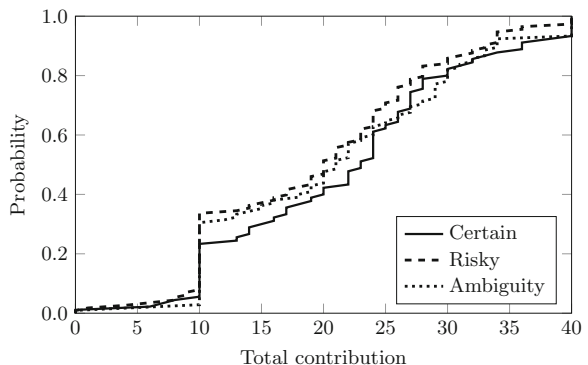
Fig. 9 Total contributions
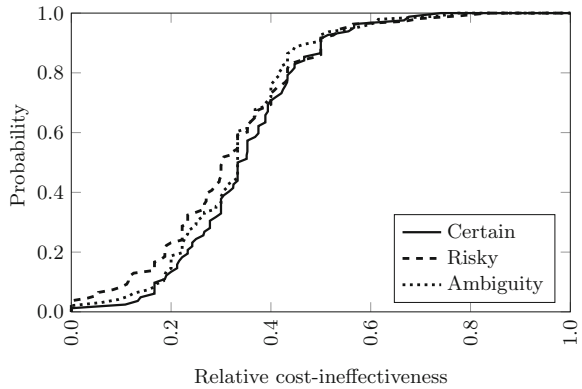
Fig. 10 Relative cost-
ineffectiveness

Table 6 Relative cost-ineffectiveness

Characteristic	Certain	Risky	Ambiguity	All
Constant	0.3094***	0.2780***	0.1290	0.2654***
Gender	0.0417	0.0511	0.0442	0.0480***
Age	0.0030	−0.0016	0.0007	0.0002
Giving type				
Pure warm-glow	0.0409	0.0329	0.1384*	0.0548
Pure altruist	−0.0459	0.0628	0.0794	0.0199
Impure altruist	0.0156	0.0252	0.1556**	0.0505
Risk attitude	−0.0155	−0.0454	−0.0509	−0.0482*
Ambiguity attitude	−0.0510	0.1845**	0.0645	0.0701*
Attention level (CRT)	−0.0530	−0.0422	−0.0336	−0.0440*
Risk literacy	−0.0791	−0.0665	0.0810	−0.0204
Observations	81	108	92	281

* 10%; ** 5%; *** 1%. Observations do not include those that do not contribute to any public good and those that contribute fully to every public good, or other questionnaire data is missing. The reference giving type is the non-donor

Cost-ineffectiveness

Figure 10 shows, for each of the three treatments, the distribution of RCI. Although, the curves are not as close to one another as in Fig. 3, a Kruskal–Wallis test does not identify a significant difference between any pair of treatments ($p = .3293$).

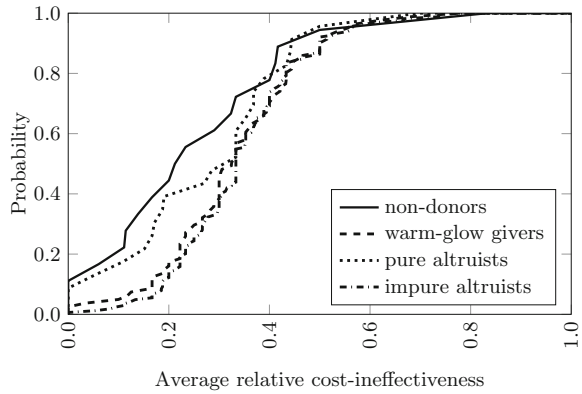**Fig. 11** Relative cost-ineffectiveness and giving type

Table 6 replicates Table 4 when considering subjects' first task only. While the core qualitative findings are essentially the same, the level of significance in some cases differs due to lower power in the between-subjects analysis.

Figure 11 replicates Fig. 4. The RCI is 0.2636 for non-donors, 0.3319 for warm-glow givers, 0.2860 for pure altruists, and 0.3423 for impure altruists. A Kruskal–Wallis test of differences in RCI across giving types is only significant at the 10% level ($p = .0773$). Subsequent Dunn's tests find the RCI to be significantly different between warm-glow givers and non-donors ($p = .0255$) and between impure altruists and non-donors ($p = .0108$). The difference between impure altruists and pure altruists is significant at the 10% level but not at the 5% level ($p = .0697$), but clearly rejects a difference for the other three pairwise comparisons ($p > .12$). These results suggest that types who are at least partially motivated by warm glow (pure warm-glow givers and impure altruists) are contributing less cost-effectively than those who are not (non-donors and pure altruists); a suggestion that can be confirmed by Mann–Whitney test comparing the RCI between these two groups of types ($p = .0140$).

Screenshots

Information and consent form

Consent Form

You are being invited to participate in this research study being conducted by Nathan W. Chan from the University of Massachusetts Amherst, Stephen Knowles and Ronald Peeters from the University of Otago, and Leonard Wolk from Vrije University Amsterdam. You were selected to participate in this study because you signed up on Prolific.

The general purpose of this research is to better understand group decision making. Participants in this study will be asked to make choices and complete tasks on a computer. If you agree to take part in this study, you will be asked to play a simple online game with two other participants, followed by a survey. We expect this study to take 45 minutes or less to complete.

You will earn 4.00 GBP for completing the study. You will receive an additional 0 to 8.28 GBP based on the choices that you and the other two participants make in the tasks, answers you give in the survey, and on chance.

You and all others taking part in this study will remain anonymous. Your Prolific ID will be used for payment purposes only, and will not be revealed to other participants and it will not be shared with third-parties. When reporting results and sharing data, we will remove individual identifying information so as to maintain confidentiality.

We believe there are no known risks associated with this research study. To the best of our ability, your answers in this study will remain anonymous. However, as with any online-related activity, the risk of a breach of confidentiality is always possible. We will minimize risks by keeping any non-anonymized data encrypted on password-protected computers.

Your participation in this study is completely voluntary and you can withdraw at any time. However, if you quit the study before completing all of the tasks, you will not earn payment for your participation. In this case, data from your decisions will be recorded up to the point that you withdraw. There are no other consequences if you choose to withdraw.

Informed consent is required for any person participating in a University research study. This study has been approved by two Institutional Review Boards for Research With Human Subjects (the University of Otago's Human Ethics Committee and the Vrije University School of Business and Economics Research Ethics Review Board; Reference codes D20/183 and SBE6/9/2020kwk350 respectively).

If you have questions about this project, or if you have a research-related problem, you may contact the researchers, Nathan W. Chan (nchan@umass.edu; +1-413-5455739), Stephen Knowles (stephen.knowles@otago.ac.nz; +64-3-4798350), Ronald Peeters (ronald.peeters@otago.ac.nz; +64-3-4798731), and Leonard Wolk (l.wolk@vu.nl; +31-20-5985262). You can contact them at any time during or after the study ends. If you have any questions concerning your rights as a research subject, you may contact the University of Otago Human Ethics Committee Administrator (Phone: +64-3-4798256; Email: gary.witte@otago.ac.nz), or the Vrije University School of Business and Economics Research Ethics Review Board (Phone: +31 205985667; Email: rerb.feweb@vu.nl)

By clicking "I agree" below you are indicating that you are at least 18 years old, have read and understood this consent form and voluntarily agree to participate in this research study. Please print a copy of this page for your records.

I agree   I do not agree

Main task

Instructions

You will be assigned to a group with two other participants. Your identity and the identities of the other two group members will remain confidential.

You and your group members will complete three independent tasks, and in each of these tasks you will make four decisions. All group members will receive exactly the same instructions as you.

Task 1	Task 2	Task 3
Decision 1A	Decision 2A	Decision 3A
Decision 1B	Decision 2B	Decision 3B
Decision 1C	Decision 2C	Decision 3C
Decision 1D	Decision 2D	Decision 3D

For each decision, each group member is given a private budget of 10 points and will decide how many of these points to contribute to the group project. Contributions can be any whole number from 0 to 10 points.

All points that the three group members contribute to the group project will be added up, and that total amount will be multiplied by a factor M. The resulting amount is the total earnings from the group project. The total earnings from the group project will be split evenly among the three group members. Your personal payoff equals your earnings from the group project, plus the part of your private budget that you did not contribute.

The payoff that you get from each decision is calculated according to the following formula:

$$Decision\ Payoff \quad = \quad \underbrace{\frac{\left(C_{you} + C_{partner1} + C_{partner2}\right) \times M}{3}}_{Earnings\ from\ group\ project} \quad + \quad \underbrace{\left(10 - C_{you}\right)}_{Private\ budget\ that\ you\ did\ not\ contribute}$$

where C_{you} is the amount that you contribute to the group project and $C_{partner1} + C_{partner2}$ is the amount contributed by your other group members.

In each task, you will make four of these decisions, where each decision may have a different value for M. Your decision screen will make very clear what this value for M is for each decision. Your earnings in a task are the sum of your earnings for each of the four decisions in this task. Therefore:

$$Task\ 1\ Earnings \quad = \quad Decision\ 1A\ Payoff + Decision\ 1B\ Payoff + Decision\ 1C\ Payoff + Decision\ 1D\ Payoff$$

$$Task\ 2\ Earnings \quad = \quad Decision\ 2A\ Payoff + Decision\ 2B\ Payoff + Decision\ 2C\ Payoff + Decision\ 2D\ Payoff$$

$$Task\ 3\ Earnings \quad = \quad Decision\ 3A\ Payoff + Decision\ 3B\ Payoff + Decision\ 3C\ Payoff + Decision\ 3D\ Payoff$$

The three tasks are very similar, but they provide different information on the values of M for each of the four decisions in the task. Make sure you read the information carefully before making your decisions.

At the end of the experiment you (and your other group members) will be paid for one of the three tasks. A computer will randomly select the task for which you will be paid. Each task is equally likely to be selected, and all group members will be paid for the same task.

All earnings are in points, and you will be paid 0.04 GBP for each point. In addition, you will receive a participation fee of 4 GBP. All relevant payment information will be provided to you on the final screen that will be shown after you have completed a questionnaire. Your earnings will only be paid out if you have completed all tasks and the questionnaire.

Click the next button once you fully understand the instructions. We expect the experiment to take approximately 35 minutes from here.

Next

Control questions

Please answer the following questions.

You will be paid your earnings from one task, which will be chosen at random.

○ False   ○ True

You will be paid for one of the four decisions in a randomly chosen task.

○ False   ○ True

For each task, you will have a total of 10 points to allocate over the four group projects.

○ False   ○ True

Suppose you contribute 6 points to the group project, partner 1 contributes 2 points to the group project and partner 2 contributes 7 to the group project. Further, suppose that the total contributions are multiplied by the factor M=2. What is your total payoff from this decision?

[   ]

[Next]

Short summary

- You will be in a group with two other participants.
- You and your partners will complete three tasks.
- In each task, you have to make four decisions.
- For each decision, there is a group project, and all group members have a budget of 10 points that they can contribute to that group project.
- Contributions to the group project by you and your partners are multiplied by a value M, and the total is divided equally among you and your partners. Hence, your payoff for a decision is given by

$$Decision\ Payoff = \underbrace{\frac{(C_{you} + C_{partner1} + C_{partner2}) \times M}{3}}_{Earnings\ from\ group\ project} + \underbrace{(10 - C_{you})}_{Private\ budget\ that\ you\ did\ not\ contribute}$$

- You will be paid for one out of the three tasks, which will be chosen at random. You will be paid for all four decisions in that task.

Task 1

You have four decisions to make. For each decision, you have a budget of 10 points.

In the boxes below, please state how many of your points you would like to contribute to the respective group project. You can contribute any whole number value from 0 to 10 points.

The value of M is different for the four group projects. **For each group project, the minimum and maximum values of M are given. Each value of M within the specified range is equally likely.** However, for each group project, you and your partners will have the same value for M.

Decision 1A. All points contributed by you and your partners to the group project will be multiplied by a value M that is between 1.20 and 1.65.
How many of your 10 points will you contribute to the group project?

```
┌─────────────────┐
│                 │
└─────────────────┘
```

Decision 1B. All points contributed by you and your partners to the group project will be multiplied by a value M that is between 1.65 and 2.10.
How many of your 10 points will you contribute to the group project?

```
┌─────────────────┐
│                 │
└─────────────────┘
```

Decision 1C. All points contributed by you and your partners to the group project will be multiplied by a value M that is between 2.10 and 2.55.
How many of your 10 points will you contribute to the group project?

```
┌─────────────────┐
│                 │
└─────────────────┘
```

Decision 1D. All points contributed by you and your partners to the group project will be multiplied by a value M that is between 2.55 and 3.00.
How many of your 10 points will you contribute to the group project?

```
┌─────────────────┐
│                 │
└─────────────────┘
```

```
┌──────┐
│ Next │
└──────┘
```

Short summary

- You will be in a group with two other participants.
- You and your partners will complete three tasks.
- In each task, you have to make four decisions.
- For each decision, there is a group project, and all group members have a budget of 10 points that they can contribute to that group project.
- Contributions to the group project by you and your partners are multiplied by a value M, and the total is divided equally among you and your partners. Hence, your payoff for a decision is given by

$$Decision\ Payoff = \underbrace{\frac{(C_{you} + C_{partner1} + C_{partner2}) \times M}{3}}_{Earnings\ from\ group\ project} + \underbrace{(10 - C_{you})}_{Private\ budget\ that\ you\ did\ not\ contribute}$$

- You will be paid for one out of the three tasks, which will be chosen at random. You will be paid for all four decisions in that task.

Task 2

You have four decisions to make. For each decision, you have a budget of 10 points.

In the boxes below, please state how many of your points you would like to contribute to the respective group project. You can contribute any whole number value from 0 to 10 points.

The value of M is different for the four group projects. **For each group project, the minimum and maximum values of M are given, and the exact value for M will be determined by a complicated random process. Therefore, we do not know the likelihood of each value of M within the range provided.** However, for each group project, you and your partners will have the same value for M.

Decision 2A. All points contributed by you and your partners to the group project will be multiplied by a value M that is between 1.20 and 1.65.
How many of your 10 points will you contribute to the group project?

Decision 2B. All points contributed by you and your partners to the group project will be multiplied by a value M that is between 1.65 and 2.10.
How many of your 10 points will you contribute to the group project?

Decision 2C. All points contributed by you and your partners to the group project will be multiplied by a value M that is between 2.10 and 2.55.
How many of your 10 points will you contribute to the group project?

Decision 2D. All points contributed by you and your partners to the group project will be multiplied by a value M that is between 2.55 and 3.00.
How many of your 10 points will you contribute to the group project?

Next

Short summary

- You will be in a group with two other participants.
- You and your partners will complete three tasks.
- In each task, you have to make four decisions.
- For each decision, there is a group project, and all group members have a budget of 10 points that they can contribute to that group project.
- Contributions to the group project by you and your partners are multiplied by a value M, and the total is divided equally among you and your partners. Hence, your payoff for a decision is given by

$$Decision\ Payoff\ =\ \underbrace{\frac{(C_{you} + C_{partner1} + C_{partner2}) \times M}{3}}_{Earnings\ from\ group\ project}\ +\ \underbrace{(10 - C_{you})}_{Private\ budget\ that\ you\ did\ not\ contribute}$$

- You will be paid for one out of the three tasks, which will be chosen at random. You will be paid for all four decisions in that task.

Task 3

You have four decisions to make. For each decision, you have a budget of 10 points.

In the boxes below, please state how many of your points you would like to contribute to the respective group project. You can contribute any whole number value from 0 to 10 points.

The value of *M* is different for the four group projects. However, for each group project, you and your partners will have the same value for *M*.

Decision 3A. All points contributed by you and your partners to the group project will be multiplied by the value $M=1.425$. How many of your 10 points will you contribute to the group project?

Decision 3B. All points contributed by you and your partners to the group project will be multiplied by the value $M=1.875$. How many of your 10 points will you contribute to the group project?

Decision 3C. All points contributed by you and your partners to the group project will be multiplied by the value $M=2.325$. How many of your 10 points will you contribute to the group project?

Decision 3D. All points contributed by you and your partners to the group project will be multiplied by the value $M=2.775$. How many of your 10 points will you contribute to the group project?

Next

Short summary

- You will be in a group with two other participants.
- You and your partners will complete three tasks.
- In each task, you have to make four decisions.
- For each decision, there is a group project, and all group members have a budget of 10 points that they can contribute to that group project.
- Contributions to the group project by you and your partners are multiplied by a value M, and the total is divided equally among you and your partners. Hence, your payoff for a decision is given by

$$Decision\ Payoff\ =\ \underbrace{\frac{\left(C_{you}+C_{partner1}+C_{partner2}\right)\times M}{3}}_{Earnings\ from\ group\ project}\ +\ \underbrace{\left(10-C_{you}\right)}_{Private\ budget\ that\ you\ did\ not\ contribute}$$

- You will be paid for one out of the three tasks, which will be chosen at random. You will be paid for all four decisions in that task.

Questionnaire

Questionnaire

You have now completed the main tasks of this experiment.

We now ask that you complete a short questionnaire.

The questionnaire consists of six parts.

Parts 1, 3, 5 and 6 are each presented on one screen; Part 2 is presented over three screens; and Part 4 over four screens. Hence, you will see ten screens in total.

In Parts 2 through 6 you can earn additional points, and hence make additional earnings.

While there is no fixed amount of time given for the first four parts, the last two parts will have a time limit. You will be given 1 minute to complete Part 5 and 2 minutes to complete Part 6.

We expect this questionnaire to take about 15 minutes in total.

Next

Demographics

Questionnaire: Part 1 of 6

What is your gender?
- Male
- Female
- Diverse
- Prefer not to say

What is your age? (Type a 0 if you do not want to say):

Next

Giving type

Questionnaire - Part 2 of 6 (1/3)

Which of the following five charity organisations do you support the most?

○ Oxfam
○ Red Cross
○ Save the Children
○ WWF
○ Doctors without Borders

Charity	Brief summary
Oxfam	Invests privately raised funds and technical expertise in local organizations around the world that hold promise in their efforts to help the poor move out of poverty; committed to long term relationships in search of lasting solutions to hunger, poverty, and social inequities.
Red Cross	Offers blood donation information and services, disaster relief, many helpful educational classes, as well as HIV/AIDS support groups.
Save the Children	Promotes children's rights, provides relief and helps support children in developing countries.
WWF	World Wildlife Fund addresses global environmental issues.
Doctors without Borders	Doctors and nurses volunteer to provide urgent medical care in some 70 countries to civilian victims of war and disasters, regardless of race, religion, or politics.

[Next]

Questionnaire - Part 2 of 6 (2/3)

You are given 20 points. You can choose to donate any amount from 0 to the entire 20 points to *Doctors without Borders*. We will also make a 20 points donation to *Doctors without Borders*, but will reduce our donation by the amount you choose to donate. Therefore, the amount *Doctors without Borders* will receive will be exactly 20 points, regardless of what you donate. However, how much of the 20 points is donated by you, and how much is donated by us, is up to you.

How much, if anything, do you choose to donate?

[_____]

[Next]

Questionnaire - Part 2 of 6 (3/3)

You just have donated 2 points, and have 18 points left. You can choose to donate any amount from 0 to the entire 18 points to *Doctors without Borders*. This time, however, we will NOT make any additional donation to *Doctors without Borders*. In total, *Doctors without Borders* will receive 20 points from the previous task plus however much you decide to donate now.

How much, if anything, do you choose to donate?

[   ]

[Next]

Risk attitude

Questionnaire - Part 3 of 6

The table below presents 11 lotteries: Lottery A through Lottery K. Each lottery has a 50% chance of paying you Payoff 1 and a 50% chance of paying you Payoff 2.

	Lottery										
	A	B	C	D	E	F	G	H	I	J	K
Payoff 1	10	12	14	16	18	20	22	24	26	28	30
Payoff 2	10	9	8	7	6	5	4	3	2	1	0
Note: Payoffs are denoted in points.											

You can choose one lottery to play. Which lottery do you choose?

○ Lottery A
○ Lottery B
○ Lottery C
○ Lottery D
○ Lottery E
○ Lottery F
○ Lottery G
○ Lottery H
○ Lottery I
○ Lottery J
○ Lottery K

[Next]

Ambiguity attitude

# Questionnaire - Part 4 of 6 (1/4)

Instructions

You will have the opportunity to earn 20 points in this task. Please read the following instructions carefully to understand the task.

There are two urns.

- **Urn X** is filled with 100 balls. Each ball is either green, red, or blue. However, it is not known how many balls of each colour are in this urn.

- **Urn Y** is filled with 101 balls. The balls are numbered from 0 to 100. There is a single ball for each number.

Urn X   Urn Y
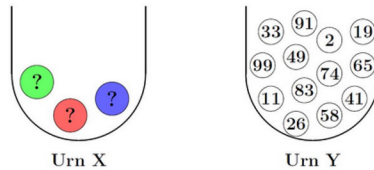
You will be given one of six possible cards. Cards are either one colour (green, red, or blue) or two colours (green/red, blue/green, or red/blue).
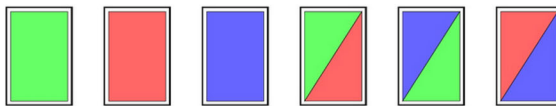
The six possible cards

You are asked to **provide a number (N)** between 0 and 100.

After you have provided your number, the computer will draw a ball from Urn Y. Then, there are two possible outcomes depending on the number (N) that you provided and the number (#) on the ball drawn from Urn Y:

- **Outcome 1:** The number (#) on the ball drawn from Urn Y is <u>lower than</u> the number (N) you provided. *In this case, your card will determine whether you win or not.* The computer will draw a coloured ball from Urn X. You will win 20 points if the colour of the ball drawn from Urn X matches the colour (or one of the two colours) on your card. Otherwise, you will earn nothing.

- **Outcome 2:** The number (#) on the ball drawn from Urn Y is <u>larger than or equal to</u> the number (N) you provided. *In this case, neither your card nor Urn X is relevant for whether you win or not.* The computer will draw a second ball from Urn Y. You will win if the number on this ball is less than the number (#) on the first ball. Therefore, the percentage chance that you will win the 20 points is equal to the number (#) on the ball that was first drawn from Urn Y. Note that this chance to win (#) is larger than or equal to the number you provided (N).

In this task, you will make this decision twice. The same two urns will be used for both decisions, and all balls will be returned to the urns between decisions. The only thing that differs between the two decisions is the colour(s) of the card that you receive. You will be paid for one of the two decisions. The computer will randomly select the decision for which you will be paid; each decision is equally likely to be selected.

Next

Short summary

- Urn X is filled with 100 balls. Each ball is either green, red, or blue, but it is not known how many balls of each colour are in this urn.
- Urn Y is filled with 101 balls. The balls are numbered from 0 to 100.

Questionnaire - Part 4 of 6 (2/4)

Please answer the following questions.

The higher your choice of (N), the more likely that your winnings will be determined by the colour of your card and the colour of the ball drawn from Urn X.

○ False   ○ True

Suppose the number (N) you provide is 63 and the ball drawn from Urn Y has number 15 on it. The chance that you will win 20 points is for sure equal to 15%.

○ False   ○ True

Suppose the number (N) you provide is 15 and the ball drawn from Urn Y has number 63 on it. The chance that you will win 20 points is equal to 63%.

○ False   ○ True

[Next]

Short summary

- Urn X is filled with 100 balls. Each ball is either green, red, or blue, but it is not known how many balls of each colour are in this urn.
- Urn Y is filled with 101 balls. The balls are numbered from 0 to 100.
- There are two possible outcomes depending on the number (N) that you provided and the number (#) on the ball drawn from Urn Y:
  - Outcome 1: The number (#) on the ball drawn from Urn Y is lower than the number (N) you provided. You will win 20 points if the randomly drawn ball from Urn X matches the colour (or one of the two colours) on your card.
  - Outcome 2: The number (#) on the ball drawn from Urn Y is larger than or equal to the number (N) you provided. The percentage chance that you will win the 20 points is equal to the number (#) on the ball that was drawn from Urn Y.

Questionnaire - Part 4 of 6 (3/4)

Decision 1:

You have a **green** card.

What number (N) do you choose?

[   ]

[Next]

Short summary

- Urn X is filled with 100 balls. Each ball is either green, red, or blue, but it is not known how many balls of each colour are in this urn.
- Urn Y is filled with 101 balls. The balls are numbered from 0 to 100.
- There are two possible outcomes depending on the number (N) that you provided and the number (#) on the ball drawn from Urn Y:
  - Outcome 1: The number (#) on the ball drawn from Urn Y is <u>lower than</u> the number (N) you provided. You will win 20 points if the randomly drawn ball from Urn X matches the colour (or one of the two colours) on your card.
  - Outcome 2: The number (#) on the ball drawn from Urn Y is <u>larger than or equal to</u> the number (N) you provided. The percentage chance that you will win the 20 points is equal to the number (#) on the ball that was drawn from Urn Y.

Questionnaire - Part 4 of 6 (4/4)

Decision 2:

You have a **red and blue** card.

What number (N) do you choose?

[   ]

Next

Short summary

- Urn X is filled with 100 balls. Each ball is either green, red, or blue, but it is not known how many balls of each colour are in this urn.
- Urn Y is filled with 101 balls. The balls are numbered from 0 to 100.
- There are two possible outcomes depending on the number (N) that you provided and the number (#) on the ball drawn from Urn Y:
    - Outcome 1: The number (#) on the ball drawn from Urn Y is <u>lower than</u> the number (N) you provided. You will win 20 points if the randomly drawn ball from Urn X matches the colour (or one of the two colours) on your card.
    - Outcome 2: The number (#) on the ball drawn from Urn Y is <u>larger than or equal to</u> the number (N) you provided. The percentage chance that you will win the 20 points is equal to the number (#) on the ball that was drawn from Urn Y.

Attention level and risk literacy

Questionnaire - Part 5 of 6

Time left to complete the three questions: **1:29**

Answer the following three questions. You will receive 5 points for each correct answer.

A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

○ 1 cent
○ 5 cents
○ 9 cents
○ 10 cents

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

○ 5 minutes
○ 20 minutes
○ 100 minutes
○ 500 minutes

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

○ 12 days
○ 24 days
○ 36 days
○ 47 days

Next

Questionnaire - Part 6 of 6

Time left to complete the four questions: **2:29**

Answer the following four questions. You will receive 5 points for each correct answer.

Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?

○ 5 out of 50 throws
○ 25 out of 50 throws
○ 30 out of 50 throws
○ None of the above

Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a random drawn man is a member of the choir?

○ 10%
○ 25%
○ 40%
○ None of the above

Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws, about how many times would the die show the number 6?

○ 20 out of 70 throws
○ 23 out of 70 throws
○ 35 out of 70 throws
○ None of the above

In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?

○ 4%
○ 20%
○ 50%
○ None of the above

Next

Waiting and feedback screens

Please wait ...

Please wait while your other group members finalize their decisions. As soon as everyone is ready, you will automatically be redirected to the next screen. This can take a few minutes.

If you think that you have waited on this screen for too long, you can try to **refresh** the page by either pressing the refresh button in your web-browser or the F5 button on your keyboard.

If you are still on this page after refreshing, then we are still waiting for your other group members to finalize their decisions.

Feedback

On the following screen you will receive feedback from your choices in the individual and group tasks. Your final payoff will also be displayed. Click on the button below to proceed.

Next

Feedback screen

Below we detail the points that you have collected during your participation.

Group task

You participated in three group tasks. The task selected for payment is: **Task 3**

The table below shows a breakdown of your payoff for each decision within the selected group task.

Decision	You kept	Total group contribution	M	Your share of group points	Your payoff
A	9.00 points	3.00 points	1.425	1.43 points	10.43 points
B	9.00 points	3.00 points	1.875	1.88 points	10.88 points
C	9.00 points	3.00 points	2.325	2.33 points	11.33 points
D	9.00 points	3.00 points	2.775	2.78 points	11.78 points

Your total payoff for all group tasks equals **44.40 points**.

Questionnaire

Item	Your payoff
Amount not donated to charity	15.00 points
Lottery outcome	4.00 points
Urns and the colored cards	0.00 points
Calculation tasks	0.00 points

Your total payoff for the individual tasks equals **19.00 points**.

Your total earnings

Your total payoff for all group and individual tasks together equals **63.40 points**. This is equal to **2.54 British Pounds** and will be paid in addition to the 4.00 GBP you receive for completing the study.

Thank you for your participation. **Click 'Next' to finalize your participation.**

Next

Data availability All data and codes used in this study are available on the Open Science Framework (OSF; https://doi.org/10.17605/osf.io/9br85) and accessible via https://osf.io/9br85/.

Declaration

Conflict of interest None.

References

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy, 97*(6), 1447–1458.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal, 100*(401), 464–477.

Andreoni, J. (1993). An experimental test of the public-goods crowding-out hypothesis. *American Economic Review, 83*(5), 1317–1327.

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics, 21*(1), 99–131.

Baillon, A., Schlesinger, H., & van de Kuilen, G. (2018). Measuring higher order ambiguity preferences. *Experimental Economics, 21*(2), 233–256.

Berman, J. Z., Barasch, A., Levine, E. E., & Small, D. A. (2018). Impediments to effective altruism: The role of subjective preferences in charitable giving. *Psychological Science, 29*(5), 834–844.

Bernasconi, M., Corazzini, L., Kube, S., & Maréchal, M. A. (2009). "Two are Better than One!: Individuals' contributions to "Unpacked" Public Goods". *Economics Letters, 104*(1), 31–33.

Björk, L., Kocher, M., Martinsson, P., & Khanh, P. N. (2016). "Cooperation under Risk and Ambiguity," Working Paper, Department of Economics, University of Gothenburg.

Blackwell, C., & McKee, M. (2003). Only for my own neighborhood?: Preferences and voluntary provision of local and global public goods. *Journal of Economic Behavior & Organization, 52*(1), 115–131.

Bramoullé, Y., & Kranton, R. (2007). Public goods in networks. *Journal of Economic Theory, 135*(1), 478–494.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29*(6), 2156–2160.

Chan, N. W., Knowles, S., Peeters, R., & Wolk, L. (2023). "On generosity in public good and charitable dictator games," Working Paper.

Chan, N. W., & Wolk, L. (2020). Cost-effective giving with multiple public goods. *Journal of Economic Behavior & Organization, 173*, 130–145.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments. *Journal of Behavioral and Experimental Finance, 9*, 88–97.

Cherry, T. L., & Dickinson, D. L. (2008). Chapter 9, Voluntary contributions with multiple public goods. In T. L. Cherry, S. Kroll, & J. F. Shogren (Eds.), *Environmental economics, experimental methods* (pp. 184–193). Routledge.

Clark, J., Garces-Ozanne, A., & Knowles, S. (2018). Emphasising the problem or the solution in charitable fundraising for international development. *The Journal of Development Studies, 54*(6), 1082–1094.

Cokely, E. T., Galesic, M., Schutz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7*(1), 25–47.

Corazzini, L., Cotton, C., & Valbonesi, P. (2015). Donor coordination in project funding: Evidence from a threshold public goods experiment. *Journal of Public Economics, 128*, 16–29.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's mechanical turk as a tool for experimental behavioral research. *PLOS One 03, 8*(3), 1–18.

Crumpler, H., & Grossman, P. J. (2008). An experimental test of warm glow giving. *Journal of Public Economics, 92*(5), 1011–1021.

Dickinson, D. L. (1998). The Voluntary Contributions Mechanism with Uncertain Group Payoffs. *Journal of Economic Behavior & Organization, 35*(4), 517–533.

Drouvelis, M., & Grosskopf, B. (2016). The effects of induced emotions on pro-social behaviour. *Journal of Public Economics, 134*, 1–8.

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior, 23*(4), 281–295.

Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization, 68*(1), 1–17.

Fielding, D., Knowles, S., & Peeters, R. (2022). In search of competitive givers. *Southern Economic Journal, 88*(4), 1517–1548.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42.

Freundt, J., & Lange, A. (2021). On the voluntary provision of public goods under risk. *Journal of Behavioral and Experimental Economics,93*, 101727.

Gandullia, L., Lezzi, E., & Parciasepe, P. (2020). Replication with MTurk of the experimental design by Gangadharan, Grossman, Jones & Leister (2018): Charitable giving across donor types. *Journal of Economic Psychology, 78*, 102268.

Gangadharan, L., & Nemes, V. (2009). Experimental analysis of risk and uncertainty in provisioning private and public goods. *Economic Inquiry, 47*(1), 146–164.

Gangadharan, L., Grossman, P. J., & Xue, N. (2023). Using willingness to pay to measure the strength of altruistic motives. *Economics Letters, 226*, 111073.

Gangadharan, L., Grossman, P. J., Jones, K., & Leister, C. M. (2018). Paternalistic giving: Restricting recipient choice. *Journal of Economic Behavior & Organization, 151*, 143–170.

Genç, M., Knowles, S., & Sullivan, T. (2021). In search of effective altruists. *Applied Economics,, 53*(7), 805–819.

Goeree, J. K., Holt, C. A., & Laury, S. K. (2002). Private costs and public benefits: Unraveling the effects of altruism and noisy behavior. *Journal of Public Economics, 83*(2), 255–276.

Granulo, A., Fuchs, C., & Puntoni, S. (2019). Psychological reactions to human versus robotic job replacement. *Nature Human Behaviour, 3*(10), 1062–1069.

Gupta, N., Rigotti, L., & Wilson, A. (2021). 'The experimenters' dilemma: Inferential preferences over populations. arxiv:2107.05064.

Hafner, R., Elmes, D., Read, D., & White, M. P. (2019). 'Exploring the role of normative, financial and environmental information in promoting uptake of energy efficient technologies. *Journal of Environmental Psychology, 63*, 26–35.

Karlan, D., & Wood, D. H. (2017). The effect of effectiveness: Donor response to aid effectiveness in a direct mail fundraising experiment. *Journal of Behavioral and Experimental Economics, 66*, 1–8. Experiments in Charitable Giving.

Levati, M. V., & Morone, A. (2013). 'Voluntary contributions with risky and uncertain marginal returns: The importance of the parameter values. *Journal of Public Economic Theory, 15*(5), 736–744.

List, J. A. (2011). The market for charitable giving. *Journal of Economic Perspectives, 25*(2), 157–80.

MacAskill, W. (2016). *Doing good better: How effective altruism can help you make a difference*. Penguin Random House.

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

Metzger, L., & Günther, I. (2019). Making an impact? The relevance of information on aid effectiveness for charitable giving. A laboratory experiment. *Journal of Development Economics, 136*, 18–33.

Null, C. (2011). Warm glow, information, and inefficient charitable giving. *Journal of Public Economics, 95*(5), 455–465. Charitable Giving and Fundraising Special Issue.

Ottoni-Wilhelm, M., Vesterlund, L., & Xie, H. (2017). Why do people give? Testing pure and impure altruism. *American Economic Review, 107*(11), 3617–33.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163.

Ribar, D. C., & Wilhelm, M. O. (2002). Altruistic and joy-of-giving motivations in charitable behavior. *Journal of Political Economy, 110*(2), 425–457.

Richefort, L. (2018). Warm-glow giving in networks with multiple public goods. *International Journal of Game Theory, 47*(4), 1211–1238.

Singer, P. (2015). *The most good you can do*. Yale University Press.

Sirota, M., & Juanchich, M. (2018). Effect of response format on cognitive reflection: Validating a two- and four-option multiple choice question version of the cognitive reflection test. *Behavior Research Methods, 50*(6), 2511–2522.

Théroude, V., & Zylbersztejn, A. (2020). Cooperation in a risky world. *Journal of Public Economic Theory, 22*(2), 388–407.

Vesterlund, L. (2015). Chapter 2, Using experimental methods to understand why and how we give to charity. In John H. Kagel & Alvin E. Roth (Eds.), *The handbook of experimental economics (Volume 2)* (pp. 91–141). Princeton University Press.

Walker, J. M., & Halloran, M. A. (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics, 7*, 235–247.

Warr, P. G. (1982). Pareto optimal redistribution and private charity. *Journal of Public Economics, 19*(1), 131–138.

Yildirim, H. (2014). Andreoni-McGuire algorithm and the limits of warm-glow giving. *Journal of Public Economics, 114*, 101–107.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.