Dodging high impact behavior with motivated beliefs?*

Frauke Stehr

May 12, 2023

Click here for most recent version

Abstract

Although other-regarding behavior is widespread, behaviors with high impact are rarely adopted. This leaves a large potential for social benefit untapped. Using an online experiment, I test the explanatory role of impact beliefs focusing on two potential cognitive mechanisms. First, motivated impact beliefs may lead to an overestimation of impact for low cost behaviors, and an underestimation of impact for high cost behaviors. Alternatively, people may think only vaguely about impact, and rather rationalize their choices ex post. I document that subjects on average overestimate low impacts slightly and underestimate high impacts. Yet, neither higher incentives for accuracy, nor changes in the costs of impactful behavior affect beliefs, implying a limited role of motivated beliefs. Reducing scope for ex post rationalization by eliciting beliefs before donations does not affect beliefs either. It does, however, increase the likelihood that subjects maximize the impact of their donation. Thus, rather than motivated beliefs, the difficulty of integrating impact and cost information across different behaviors seems to play a role in the low adoption of high impact behaviors.

JEL codes: D91, D64, D83

*First version: September 14, 2022. This version: May 12, 2023. Contact information: Frauke Stehr, Maastricht University. E-mail: f.stehr@maastrichtuniversity.nl. Website: https://fstehr.github.io. Funding by the Graduate School of Business and Economics at Maastricht University is gratefully acknowledged. I thank especially my supervisors, Christian Seel and Peter Werner, as well as Kai Barron, Luigi Butera, Antonio Cabrales, Nickolas Gagnon, Katrin Gödker, Bertil Tungodden, Lise Vesterlund, Roberto Weber, Florian Zimmermann and seminar participants in Maastricht, Berlin, Bergen, Bologna and Zürich for helpful comments and discussions. The experiment obtained ethics approval by the Ethical Review Committee Inner City Faculties at Maastricht University (ERCIC_256_26_04_2021). The main hypotheses and analyses were pre-registered on AsPredicted.org (#93434).

1 Introduction

Other-regarding behavior is omnipresent. In 2018, for example, 49% of Americans gave to charity (Indiana University Lilly Family School of Philanthropy 2021), 30% spent time volunteering (AmeriCorps 2021), and 89% made an effort to live in an environmentally friendly manner in 2019 (Pew Research Center 2019). Yet, the specific behaviors that are adopted often have only small impact although higher impact actions are available.[1] One prominent example comes from the environmental domain, where many people focus on low-impact actions such as switching off lights or sorting waste (Dubois et al. 2019). However, even the most pro-environmental individuals rarely adopt highly impactful behaviors such as flying less or adopting a vegan diet (Diekmann and Preisendörfer 2003).[2] Similar observations can be made with donations where, for example, charities often receive a plethora of goods after disasters that are of negligible use for the recipients (e.g., Osman 2011). This leaves large untapped potential for individuals to adopt more effective behaviors.

The aim of this project is to improve our understanding of why the adoption of high-impact other-regarding behaviors is low. To this end, I conduct an online experiment to test potential underlying cognitive mechanisms that might explain these patterns. Understanding these mechanisms is crucial to be able to determine appropriate policy responses. If high-impact behaviors are rarely adopted because of biased beliefs, a correction of beliefs might be an effective tool, as long as these beliefs are not biased for self-serving reasons.[3] For a bounded rationality mechanism, decision aids for example simplifying comparisons across behaviors might be most effective. Finally, both information provision and decision aids might be ineffective at changing behavior if the low adoption of high-impact behaviors would be preference-driven.

Three observations guide my hypotheses. First, even though impact information is available in principle, the demand for it seems low even among donors.[4] As a consequence, impact beliefs are often biased, with survey results suggesting that people overestimate low impacts and underestimate high impacts (e.g., Ipsos 2021). Second, higher impact behaviors typically are also more costly to implement, both in terms of actual costs as well as opportunity costs. A traditional explanation would thus be that the costs of high-impact behaviors are prohibitively high. Yet, these price

[1] I use the term impact to capture the social benefit of an action as opposed to its efficiency (i.e., impact per cost).

[2] While also more self-oriented motives may explain pro-environmental behavior, recent evidence shows that pro-environmental behavior is well predicted by altruism and moral universalism (Lades, Laffan, and Weber 2021; Enke, Rodriguez-Padilla, and Zimmermann 2022).

[3] Several mechanisms might limit the effectiveness of information campaigns when beliefs are biased because of motivated reasoning. Not only might people then try to protect their ego by evading/ignoring the information provided, motivated memory might also lead them to asymmetrically forget bad news (Zimmermann 2020).

[4] There are various websites that estimate charities' relative effectiveness, e.g., GiveWell. Also carbon footprint calculators are meant to inform about relative impacts of different behaviors. Demand for such comparative information is, however, low. For example, only 3% of surveyed donors claimed to have done research comparing charity efficiency (Hope Consulting 2012).

differences can also serve as drivers of motivated impact beliefs for people with (self-)image concerns. Suppose people are only willing to take cheap actions, but still want to feel good about themselves. This motivation may lead subjects to *over*estimate the impact of cheap projects and *under*estimate the impact of expensive projects to justify why they are already doing enough when engaging in the cheap action. Third, when deciding which behaviors to adopt, people are not forced to think about impact. This leaves room for not thinking about impact when making the choice for a cheap behavior and rather rationalize choices *ex post* with biased impact beliefs.

Building on these observations, I leverage a donation experiment to isolate the roles of (ex ante) motivated impact beliefs and ex post rationalization. In the experiment, subjects receive an endowment they can donate to a charity that administers vitamin A supplements to children in need. I vary costs and impact levels independently across projects. Costs of donating can be either high or low and are always known from the beginning. I create variation in impact by varying the number of vitamin A supplements that a donation finances, making it easy for participants to think of the impact as quantifiable. Unlike costs, impact has to be estimated from a noisy signal. A signal is a matrix filled with two different symbols, where more pill symbols reflect more vitamin A doses financed. Subjects see these matrices for a short while and are then asked how many pill symbols the thought were in the matrix. This measure of impact beliefs is my main outcome variable of interest. In a given round, subjects always face two projects and can donate to neither, one or both of them. In the main parameter combination of interest, subjects face the incentive structure outlined above: they can donate to a high cost project with high impact, and to a low cost project with low impact. This creates a clear trade-off for subjects between the costs they have to bear and the impact their donation can create.

I introduce two treatment variations to isolate the roles of motivated beliefs and the scope for ex post rationalization in a 2x2 design. In the first treatment dimension, I vary within subject the strength of incentives for accurate beliefs from a low accuracy bonus ($LoAB$) to a high bonus ($HiAB$). When subjects trade off the utility of their self-image against the utility of money, higher incentives for correct beliefs should reduce the incentive for motivated beliefs (e.g., Bénabou and Tirole 2002). In a second treatment dimension, I change the timing of eliciting beliefs from the attention task between subject. In both conditions, subjects first see the cost of donation. In *ExPost*, they then see the two signals one after the other, after which they decide whether to donate. Having decided on their donations, I then elicit beliefs about the number of pills in the two matrices. In contrast, in *ExAnte*, I elicit beliefs *before* donation choices, thus making subjects think about impact before donating, thereby making ex post rationalization more difficult. At the same time, eliciting beliefs before donation choices makes it easier to integrate information about the cost and the impact of a project at the time of donating. Finally, I run a *NoChoice* treatment that removes the incentive to form motivated impact beliefs entirely. In this treatment, subjects are only asked to indicate their impact beliefs for the donation projects. At the same time, I make clear that they will never have to donate to these projects, thus removing the incentive for motivated impact beliefs.

As an exploratory research question, I investigate how sensitive impact beliefs and donation patterns are to changes in trade-offs. In particular, I vary the choice sets within each treatment block in a structured way, changing cost and impact levels one by one. This leaves subjects with two sets of choices where both projects differ in cost but not impact, as well as two sets of choices where both projects have the same costs but differ in impact. This variation gives me another channel to test for motivated beliefs, by comparing beliefs across different prices of donating. On top, this variation in cost and impact allows me to estimate both impact and price elasticities.

Several findings emerge from this study. First, looking at beliefs, I find that although initially biased, subjects' beliefs are robust to various treatment changes, suggesting a limited role of motivated beliefs. Subjects on average (slightly) overestimate low impact and markedly underestimate high impact levels in the treatment with low incentives for accuracy. Belief patterns are highly heterogeneous across subjects, with 22% of subjects displaying this exact pattern of overestimating low impacts and underestimating high impacts.[5] At the same time, beliefs are robust to treatment changes. Contrary to previous findings and theoretical predictions (e.g., Zimmermann 2020), a ten-fold increase in incentives does not improve belief accuracy. Removing the donation choice in *NoChoice* to take away intrinsic motives to bias beliefs does not change beliefs either. Subjects also do not seem to engage in ex post rationalization, as there is no significant difference in beliefs comparing *ExPost* and *ExAnte*. Finally, subjects' beliefs do not react significantly to changes in the price of donation. Testing for heterogeneous treatment effects, I find that beliefs are not significantly different for subjects with high and low levels of altruism.

Second, subjects are responsive in their donations to changes in impact levels and beliefs (conditional on true impact). Still, the median subject donates only to the low cost, low impact project, i.e., they donate $4 out of their $40 endowment, thereby financing 8 doses of vitamin A.[6] More altruistic subjects, as measured both by their willingness to pay for 32 doses of vitamin A, and a survey measure of altruism (Falk et al. 2016), are more likely to donate and hence create a larger impact. Subjects who donate generally have a higher impact belief than non-donors in all treatments. In particular, donors believe that the low impact project finances 1.73 more doses (21%), and that the high impact project finances 2.40 more doses (9%) compared to non-donors' beliefs. Still, donors underestimate the high-impact project on average by 2.35 doses. Consistent

[5] I classify subjects as accurate, under- or overestimators, separately for their two beliefs in the parameter combination with a trade-off between cheap, low impact and expensive, high impact, in the *LoAB, ExPost* treatment. 22% of subjects do not consistently over- or underestimate, but overestimate cheap low impact and underestimate expensive high impact.

[6] Based on recommendations by the World Health Organization (WHO), in total 9 doses should be administered to children aged 6–59 months in areas with high rates of vitamin A deficiency to reduce morbidity and mortality (WHO 2011).

with finding no change in beliefs with the higher accuracy bonus, I also do not find that donations change when increasing incentives for accurate beliefs. To test whether subjects do not react to the change in incentives at all, I exploit the panel structure of my data to explore the variations in beliefs and donations on the subject level. I find that subjects indeed change their beliefs in response to higher incentives for accuracy and that this change positively correlates with a change in donation. That is, subjects who correct their belief upwards with higher incentives for accuracy are also more likely to donate more. The reverse is also true: subjects who correct their belief for the high impact project downwards with higher incentives for accuracy are also more likely to donate less.

Third, eliciting impact beliefs before donations, thereby nudging subjects to think about impact, changes donation behavior. In particular, subjects in *ExAnte* are 6 percentage points more likely to maximize impact by donating to both projects, compared to a baseline of 15% in *ExPost*. The structured changes in trade-off combinations allow me to calculate both price and impact elasticities in both treatments. Donations are generally inelastic with respect to changes in prices and impact, that is for a 1% change in price or impact, the likelihood of a donation to that project changes by less than 1%. Interestingly, subjects react more strongly to changes in impact compared to changes in prices. Comparing elasticities across treatments, I find that while there are no differences in price elasticities, subjects react less strongly to changes in impact in the *ExAnte* treatment, with impact elasticities on average 0.24 lower than in *ExPost*. One explanation for this could be that eliciting impact beliefs before donation decisions makes it easier to aggregate impact across the different projects, thus reducing sensitivity to changes in impact for the individual project.

The results of this paper suggest that motivated cognition plays a limited role in explaining why high impact behaviors are rarely adopted. Rather, difficulty integrating information about costs and impacts of different projects seems to drive donation patterns. Having indicated one's impact beliefs before donating makes it easier to compare and aggregate impact information across projects. While this does not seem to affect beliefs, it does make it more likely that subjects donate to both projects and thus maximize their impact. By shedding light on the mechanisms underlying impact belief formation and donation behavior, the results of this paper have policy implications, for example for information design and choice architecture. Due to the limited role of motivated beliefs, the results suggest that making impact information easily available and comparable may increase the impact of adopted behaviors. Taken together, these findings highlight the important role tools like charity impact evaluators and carbon footprint calculators can play in directing other-regarding behavior towards high impact actions.

Related Literature This paper builds on and contributes to several strands of the literature. First, by directly studying the formation of impact beliefs as an explanatory mechanism, I contribute to the literature on impact (in-)sensitivity in behavior adoption and ineffective altruism

(Caviola et al. 2020; Caviola, Schubert, and Greene 2021; Genç, Knowles, and Sullivan 2021; Heeb et al. 2022; Metzger and Günther 2019). Such insensitivity to impact has, for example, been documented in Hagmann, Ho, and Loewenstein (2019). They show that absent impact information people react to decoys in their support for policies of varying impact. For example, they reduce their support for (seemingly) expensive, high impact policies (substantive taxes) when cheaper, but less impactful policies (nudges) are also part of the choice set. Providing impact information can eliminate this effect. This suggests that impact beliefs might be biased, which I explicitly test in my experiment. My finding that subjects overestimate low impact levels and underestimate high impact levels corroborates their results. The finding that eliciting impact beliefs prior to donations does not change beliefs but increases the likelihood of maximizing impact suggests that difficulty in integrating impact information can play a determining role in explaining low impact sensitivity. I thereby contribute to recent work which shows that complexity in aggregating information about different impact dimensions and that cognitive uncertainty are important drivers in under-sensitivity to impact (Toma and Bell 2022). In their paper, they experimentally investigate how laymen and policymakers aggregate different dimensions of impact of various policy proposals and how this aggregation affects funding decisions. They find that people are generally under-sensitive to impact, that is, they react less than 1:1 in their funding decisions to changes in impact. In contrast to them, I look at a setting where impact information is noisy, but relatively simple, which allows me to study in a controlled way how the costs of different behaviors affect impact belief formation.

Second, this paper contributes more generally to the literature on motivations of other-regarding behavior (Andreoni 1990; Dana, Weber, and Kuang 2007; Gino, Norton, and Weber 2016; Filiz-Ozbay and Uler 2019). In particular, I propose and test a mechanism that could bridge the gap between the literatures on excuse-driven behavior and warm-glow giving. The former has documented that people use excuses and tend to bend their beliefs and preferences to justify acting selfishly (Exley 2015; Exley and Kessler 2019). This seemingly contradicts the body of evidence that people donate even when it is inefficient to do so, often attributed to a feeling of warm-glow (Null 2011; Ottoni-Wilhelm, Vesterlund, and Xie 2017). I hypothesize that this seeming contradiction may be resolved accounting for the costs of different behaviors – people may want to find excuses to act selfishly when the behavior is too costly for them, and at the same time enjoy the warm glow benefit of relatively cheap actions. Thereby, I contribute to recent work showing theoretically that also donors may not want to learn about charities' impact, when sufficiently driven by warm-glow, to maintain optimistic beliefs that justify their giving (Niehaus 2020). While there is evidence showing that also donors like to avoid impact information (Jhunjhunwala 2021), I am to the best of my knowledge the first to explicitly test whether impact beliefs are indeed biased, and whether this is driven by the cost of donating. I find that subjects beliefs are – although biased – robust to changes in the price of donating, which suggests that accounting for costs of donating cannot overcome this seeming contradiction in (experimentally) observed donation patterns. The fact that beliefs patterns are very heterogeneous and around 22% of participants display joint over- and underestimation suggests instead that there might be a continuum of types in the population with differing weights on warm-glow and belief-based utility (excuse-driven givers).

Third, the paper contributes to the literature on motivated beliefs (Bénabou and Tirole 2002; Zimmermann 2020). More specifically it relates to the literature looking at motivated beliefs on social impact and externalities (e.g., theory: Hestermann, Le Yaouanq, and Treich 2020; experiments: Di Tella et al. 2015; Ging-Jehli, Schneider, and Weber 2020; Ahumada et al. 2022) as well as the newer literature on cognitive flexibility and ex post rationalization (Saccardo and Serra-Garcia 2022; Eyster, Li, and Ridout 2021). My finding that beliefs are robust to various changes in incentives and decision environment adds to some recent null results findings in different environments (Pace and Weele 2020; Gangadharan, Grossman, and Xue 2021a; Engelmann et al. 2022), suggesting that more research is warranted to better our understanding of when and how motivated beliefs exactly occur, and how to elicit them in more abstract experimental settings.

Finally, this paper adds to the literature on moral licensing and compensatory behavior (e.g., Gneezy, Imas, and Madarász 2014; Blanken, Ven, and Zeelenberg 2015; Maki et al. 2019; Chater and Loewenstein 2022). This literature investigates how having acted morally in one domain, may lead to acting less morally subsequently. In the literature on pro-environmental behavior this has been called behavioral spillover effect. One example from the environmental domain is Tiefenbeck et al. (2013), which documents that residents after an intervention aimed at reducing water consumption, indeed reduced their water consumption, but at the same time increased their electricity consumption. I contribute to this literature by investigating the role of impact beliefs as a mechanism facilitating compensatory behavior.

2 Experimental Design

2.1 The decision situation

Subjects play various modified dictator games with a charity recipient. In each of multiple rounds, they are shown different projects to which they can donate using their endowment of $40. Projects differ in price and impact. Before making a donation decision, prices of donating to each of the projects are always perfectly known, and could be high ($16) or low ($4). Impact, however, has to be estimated from a noisy signal, and could also be either high or low. In particular, subjects' donation finances varying numbers of doses of vitamin A supplements (8 doses for low impact projects, and 32 doses for high impact projects). Donating vitamin A supplements has the advantage of being clearly quantifiable (by the number of doses financed), which makes it easier for subjects to think about varying magnitudes of impact.[7] Subjects are informed that there are these two price levels for donations. At the same time, they are not informed that there are only two impact levels, instead they have to estimate impact for each project. Importantly, subjects cannot immediately infer a project's impact from the price, as impact and prices are varied independently across rounds.

Subjects see two projects each round, called project A and project B. They can donate to one, both or neither of these projects. This serves two purposes. First, it increases the salience of differences between projects, compared to showing the various projects one after the other. Second, not forcing subjects to donate (for example by having them choose one project they would like to donate to), allows for a positive self-signaling value of donating to the cheaper option.[8] Subjects receive an endowment of \$40 in each round. When donating, subjects receive the endowment minus the sum of prices of the projects to which they donate. On top, I donate \$1.10 $\times$ the number of vitamin A doses of the respective projects to the charity Helen Keller International.[9] When not donating at all, subjects can keep the entire endowment and no vitamin A donations are triggered. To avoid hedging, one choice is randomly drawn to determine which payments and donations are realized.

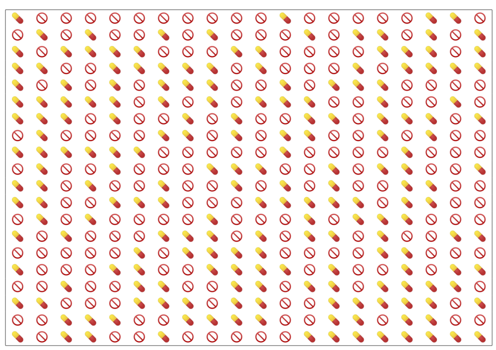


Figure 1: A signal with 200 pills

To implement a noisy signal of impact, which still makes clear that there is no inherent uncertainty about whether the project has an impact at all, I adapt an attention task previously used in the literature (Ambuehl 2017; Pace and Weele 2020; Bosch-Rosa, Gietl, and Heinemann 2021). A signal is a 20×20 matrix containing two different symbols in varying proportions – a pill and a prohibited

[7]Subjects are informed that vitamin A deficiency is responsible for death in children and that the WHO recommends vitamin A supplements to children in regions where vitamin A deficiency is a public health problem.

[8]In models of image concerns, agents choose actions to signal their (pure) intentions to themselves or others (Bénabou and Tirole 2002). When subjects would be forced to donate to either of the projects, there is a pooling equilibrium on the cheap donations. Specifically, people with a low willingness to pay for donating will choose the same action as the selfish people who would have preferred not to donate at all. This makes it harder for weakly altruistic subjects with image concerns to signal (to themselves) that they are actively doing something altruistic, as a selfish type would have picked the same option.

[9]\$1.10 is the expected cost per dose of vitamin A for this charity (GiveWell 2021). In expectation, subjects' donation thereby finances the administration of the respective number of vitamin A doses.

sign (Figure 1). More pills in a matrix reflect more impact, with ten pills reflecting one dose of vitamin A supplements. For each impact level, I create multiple matrices with different random pill patterns. This makes it more difficult for subjects to recognize across rounds that impact, in fact, has only two levels. Impact levels are chosen such that high impact signals (320 pill emojis) would be – in principle – equally difficult to interpret as low impact signals (80 pill emojis), as one could estimate the number of prohibited signs and subtract it from 400. To increase the chance of finding motivated beliefs compared to a neutral frame, I use contextual framing and symbols in the matrices to improve participants' understanding of the decision context and the relevance of the matrices for donations (Alekseev, Charness, and Gneezy 2017).

To avoid counting and to limit the extent to which subjects endogenously differ in the amount of effort spent looking at the matrix, subjects see each matrix for 7 seconds only. In the baseline treatment (*ExPost*), I ask for donation choices directly after subjects see the signals to allow for ex post rationalization. After submitting their donation choices, subjects indicate their belief as well as their confidence in their estimate.[10] Accuracy in beliefs is incentivized in a simple way, following recent evidence that for online samples easier belief elicitation procedures may induce less biases (Danz, Vesterlund, and Wilson 2020; Burdea and Woon 2022): for any belief not more than 10 away from the true number, subjects could receive a bonus payment. To elicit subjects' confidence in their beliefs, I use the cognitive uncertainty slider introduced in Enke and Graeber (2021), which indicates an implied confidence interval when moved.

I elicit all measures on one screen to make the relevance of the belief task for the donation decision salient to subjects.[11] In particular, when seeing the signals, subjects also see already the donation prompt albeit deactivated (Figure 12 in the appendix). It is made visually salient which field is currently active to be filled out; deactivated fields are grayed out. The timing of each screen is as follows:

1. An alert in the foreground asking participants to stay attentive (3 seconds)
2. matrix A displayed (7 seconds, then button click)
3. matrix B displayed (7 seconds, then button click)
4. the donation fields are activated (until button click)
5. the belief field for A is activated ($\leq$ 20 seconds);
   the cognitive uncertainty slider for belief A is activated (until button click)
6. the belief field for B is activated ($\leq$ 20 seconds);
   the cognitive uncertainty slider for belief B is activated (until button click)

[10]To avoid that subjects take a screenshot and count, I limit the answer time to 20 seconds. A pilot test indicated that this was enough time to fill in both those measures.

[11]Screenshots of all instructions as well as the decision screen can be found in the appendix.

Subjects always first see the project on the left, but I randomize which project of a specific parameter combination is shown first. Before starting with the payment relevant part, subjects participated in a trial round to familiarize themselves with the decision environment. At the end of the experiment, I elicit subjects' willingness to pay for various numbers of vitamin A doses using multiple price lists (MPLs). The price lists enforce single switching, but not weakly increasing willingness to pay. I also collect basic demographics (age, gender, household-income level, education) as well as survey measures of altruism and warm-glow (Falk et al. 2016; Carpenter 2021).

2.2 Treatments

To isolate the effect of motivated beliefs within subject, I vary the bonus subjects receive for stating accurate beliefs, following Zimmermann (2020). If subjects trade-off their consumption-based utility with a belief-based utility of self-image, higher bonuses for accurate beliefs may tilt the incentives in favor of consumption-based utility and thus reduce motivated beliefs (Bénabou and Tirole 2002). In the low accuracy bonus treatment (*LoAB*), subjects receive a bonus of \$2 when their indicated belief is not more than 10 away from the true number of pills in a matrix. In the high accuracy bonus treatment (*HiAB*), subjects receive a bonus of \$20 instead. This treatment variation is applied within subject for all between subject treatments described below. To be able to control for order effects, I randomize the order of these treatments on the subject level.

To isolate the effect of ex post rationalization, I vary the timing of the belief elicitation task between subject. In particular, I change the point of time at which subjects are asked to indicate their belief. In *ExPost*, as described above, subjects first see the signal for project A, then the one for project B, and are then asked for their donation choice, before moving on to indicate their two respective beliefs. In *ExAnte*, on the other hand, subjects first see the signal of project A, indicate their belief, see the signal of project B, indicate their belief and only then give their donation choice, while still seeing the beliefs they had given earlier (see Figure 13 in the appendix). Having to indicate impact beliefs before choosing where to donate may induce subjects to think about impact before making their donation choices. Additionally, having already indicated one's beliefs makes it easier to integrate impact information upon donating, while also making ex post rationalization more difficult.

2.3 Expected Behavior

The main research question in this paper is whether people overestimate low and underestimate high impact levels of their other-regarding behavior for self-serving reasons. In particular, when costs and impact are positively correlated, an agent with low (but positive) valuation for other-regarding behavior (say charitable giving, in the context of this experiment) might want to engage in self-deception as a justification for not donating to an expensive project. One way to justify such a decision – when beliefs about costs of donating are exogenously fixed – is by changing one's impact belief. In particular, underestimating the impact of the expensive project makes it look less attractive, and hence can serve as a further justification/excuse for not donating to this project. Such excuse-seeking behavior has been documented before (e.g., Dana, Weber, and Kuang 2007; Exley 2015, 2020). I extend the findings from the previous literature by hypothesizing (and testing) that such excuse driven behavior is driven by high costs of donating.

At the same time, an agent with a low, but positive valuation for other-regarding behavior may find it difficult to derive utility from a donation to a cheap project with a price below their valuation, as this project's impact level may be quite low.[12] In this case, the agent may not want to learn about true impact, and prefer to (effectively) overestimate impact, to justify her donation choice (Niehaus 2020). This allows her to keep on donating and thereby receiving a warm-glow bonus.

I design the experimental treatment conditions to isolate two mechanisms which may yield such biased beliefs: ex ante motivated impact beliefs and ex post rationalization of donations' impact. In particular, I hypothesize that treatments which allow for motivated beliefs (i.e., the *LoAB* condition) lead to the overestimation of low impact levels, while they lead to an underestimation of high impact levels. That is, comparing beliefs in the treatment with low bonus for accurate beliefs (*LoAB*) to the treatment with high bonus for accurate beliefs (*HiAB*) would yield one test of motivated impact belief formation. An additional, different test was pre-registered as exploratory analysis: comparing beliefs for the same impact level across different costs of donating.

Conditional on such beliefs (biased in the way described above), I pre-registered to test whether beliefs are biased *instrumentally* to justify donation decisions. That is, if beliefs change in the treatment which makes holding motivated beliefs more costly (*HiAB*), do donation patterns also change? In particular, if subjects would have more accurate beliefs in *HiAB*, I expect to see more frequent donations to the expensive, high impact project, whose impact was underestimated in the *LoAB* treatment. In a similar vein, I hypothesize to see fewer donations to the cheap, low impact project, whose impact I hypothesize be overestimated in the *LoAB* treatment.

Finally, comparing beliefs between the *ExPost* and the *ExAnte* treatment allows me to see whether people make use of the increased moral wiggle room when indicating their impact beliefs only once donation decisions have been made. If subjects are willing to engage in such ex post rationalization, I expect to see more biased beliefs in the *ExPost* compared to the *ExAnte* treatment.

[12] Such utility can be derived either by warm-glow of giving, or self-image/social-image concerns. For both, at least a weakly positive impact belief is necessary.

2.4 Experimental Procedure

The experiment was run in April 2022 on Prolific, after some parameter pilot tests. All hypotheses were pre-registered on AsPredicted under registry entry #93434. As pre-registered, I collected a sample of 900 US-based subjects (600 in *ExPost*, 300 in *ExAnte*), over-sampling the *ExPost* treatment due to an expected increase in variance of beliefs. This sample size would allow me to detect mean differences in beliefs of 8.6 (=0.16 SD) in the *ExAnte* condition at 5% significance with a power of 80% (based on pilot standard deviations). The sample was gender-balanced.

Instructions were delivered in separate bunches with comprehension questions in between. Subjects had to correctly answer all comprehension questions before being admitted to the experiment.[13] All subjects received a participation fee of $3.50. Additionally, I randomly selected 20 subjects who would be paid based on one of their decisions which was independently randomly drawn. To avoid hedging between rounds, either one of their beliefs, or a donation decision, or one of the multiple price lists that were used for the WTP elicitation could be drawn to be implemented for payment. Subjects took on average 20.33 minutes to complete the experiment.

3 Results

3.1 (Motivated) Impact Beliefs

In the following, I analyze beliefs and donations when subjects face a trade-off between a low cost, low impact and a high cost, high impact donation. I start the analysis by looking at impact beliefs in the *ExPost* treatment. Subjects on average *over*estimate the number of pills in the matrix of the low cost/low impact project by 6.76 in the treatment with low accuracy bonus, *LoAB* ($p = 0.0086$, two-sided t-test of difference from 80). At the same time, subjects *under*estimate the number of pills of the high cost/high impact project by 33.39 on average ($p < 0.001$, two-sided t-test of difference from 320).[14] Around 22% of subjects display this exact pattern of joint over- and underestimation.[15] Note that overestimation is of a much smaller magnitude than underestimation. As can be seen from Figure 2 beliefs for the high impact project are not only less accurate on average, but also more noisy in general ($p < 0.001$, F-test of equal variances). This might be driven by imprecise

[13] Subjects always had access to the instructions when answering comprehension questions. They were given multiple attempts to answer correctly, but to be admitted to the experiment they were not allowed to answer wrongly more than two times.

[14] All statistical tests reported are two-sided unless otherwise indicated.

[15] From the density plot in Figure 2, it seems that average beliefs are affected by the long tails in the distributions. When looking at the median of impact beliefs instead, subjects underestimate low impact (median belief: 75 in *LoAB*, 70 in *HiAB*). For the high impact project, however, also the median subject underestimates impact (median belief: 310 in both incentive conditions).

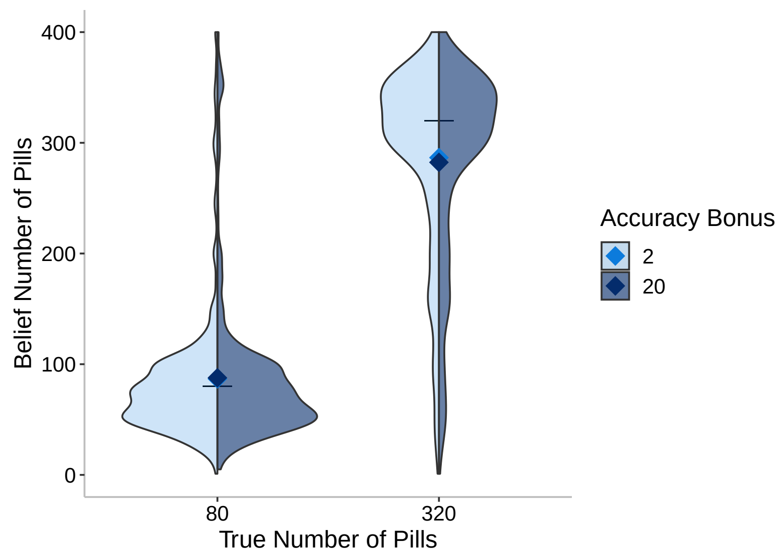


Figure 2: Distribution of impact beliefs in the two accuracy bonus conditions, for the *ExPost* treatment. Diamonds show averages compared to true value (horizontal bars). Shaded areas represent densities (light grey=*LoAB*, dark grey=*HiAB*). Beliefs indicate perceived impact of the low cost, low impact project (on the left) and the high cost, high impact project (on the right).

perceptual judgment, which makes it more difficult to process larger numbers leading to stronger central tendency (Woodford 2020; Xiang et al. 2021).

To test whether beliefs are biased for self-serving reasons, I compare beliefs in the low and high accuracy bonus treatment. The darker shaded distributions in Figure 2 indicate that beliefs are very similarly distributed across treatments. This is confirmed by the pre-registered paired t-tests ($p = 0.78$ for the low cost/impact project, $p = 0.42$ for the high cost/impact project). Regression results accounting for subject differences also confirm this (Models (1) and (4), Table 1). Previous findings in the literature (e.g., Zimmermann 2020) and theoretical results (Bénabou and Tirole 2002; Brunnermeier and Parker 2005) would suggest that there is a fundamental trade-off between self-deceptive beliefs which serve to maximize belief-based utility, whereas outcome-based utility is better served by accurate beliefs. Increasing incentives for accurate beliefs would thus change the trade-off between these two utilities in favor of more outcome-based utility leading to more accurate beliefs. The fact that a 10-fold increase in incentives does not change beliefs suggests that motivated beliefs play only a limited role in the setting of my experiment. While the effect of higher incentives could be bounded by a desire for consistency due to the within subject nature of the treatments, this is unlikely. First, subjects generally do not appear to have constant beliefs across rounds (Figure 8 in the Appendix). As a second robustness check, I compare beliefs between subjects across incentive schemes in the first part only (i.e., before being exposed to the other incentive treatment). Also here beliefs do not differ significantly between incentive schemes (p =

	low cost/impact			high cost/impact		
	(1)	(2)	(3)	(4)	(5)	(6)
HiAB	0.99	1.71	1.71	−4.17	−8.98	−8.99
	(2.83)	(3.77)	(3.77)	(3.65)	(4.90)	(4.90)
below_med_wtp		5.26	7.46		1.14	1.53
		(5.17)	(5.11)		(6.93)	(6.99)
HiAB x below_med_wtp		−1.51	−1.47		10.25	10.18
		(5.68)	(5.69)		(7.32)	(7.34)
Constant	85.84***	85.48***	47.53*	308.59***	310.99***	340.69***
	(19.95)	(19.86)	(22.06)	(9.94)	(10.13)	(17.35)
# obs.	1209	1209	1207	1210	1210	1208
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	Yes	No	No	Yes
# subj.	607	607	606	607	607	606

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Linear RE-regressions including session FE with impact belief in the *ExPost* treatment as dependent variable. Clustering-robust standard errors in parentheses. The variable below_med_wtp is a dummy variable taking value 1 if a subject's willingness to pay for 32 vitamin A doses was below median. *HiAB* is a dummy variable, taking value 1 when the bonus payment for accurate beliefs was high.

0.87 for the low cost/impact project, p = 0.33 for the high cost/impact project; two sample t-test).

Another potential explanation for the absence of an incentive-induced effect on belief accuracy might be heterogeneous treatment effects. As pre-registered, I therefore look at whether the effect of *HiAB* is stronger for subjects with low levels of altruism. These subjects might be less intrinsically motivated to have accurate beliefs and hence react more strongly to a change in incentives (as suggested e.g., in Gangadharan, Grossman, and Xue (2021b)). I test this by looking at subjects who have below median willingness to pay for 32 doses of vitamin A elicited in a multiple price list at the end of the experiment, as specified in the pre-registration. As the regressions in Table 1 show, however, neither do subjects with a below median willingness to pay for impact have a significantly different belief from subjects with a high willingness to pay, nor do they react differently to changes in incentives (models (2) and (5)). If anything, it seems that subjects with above median willingness to pay are more responsive to changes in accuracy incentives (models (5) and (6)), but these subjects downward adjust their belief for the high impact treatment. Adding demographic controls (models (3) and (6)) does not change the marginal effects of the treatment.

Several other reasons could explain this robustness of beliefs in response to an increase in incentives. It could be that subjects do not perceive the trade-off between belief-based and outcome-based utility to be strong as they will never learn (let alone 'feel') how effective their donations actually were. Findings from Jhunjhunwala (2021) suggest for example, that regret aversion plays an important role in information avoidance on the impact of donations. Informing subjects a priori that they would eventually learn how effective the charities in the experiment were led to significantly more search and more efficient donation choices. Since subjects in my experiment know that they will never learn the true impact levels, the increase in incentives for accurate beliefs in my experiment may not be enough to reduce the intrinsic incentive for motivated beliefs. Another reason might be that subjects in fact do react to higher incentives, just not by adjusting their beliefs. Engelmann et al. (2022) show that while higher incentives do not reduce wishful thinking in their experiment, they do increase effort as measured by self-reported concentration and response times. In my experiment, I limit response times to avoid that subjects online take screenshots of the matrices and then count. One indicator of more effort produced might be the reported certainty in their belief. However, I do not find that higher incentives for accurate beliefs robustly increase certainty. In fact, there is no effect on reported certainty for low impact projects, and only a small effect for high impact projects (Table 6 in the appendix). Importantly, this effect is not large enough to affect belief accuracy (Table 7 in the appendix).

3.2 Donations

I next turn to donation behavior. A majority of subjects donate at least to the cheap, low impact project, with 16% donating to the cheap, low impact project only, 36% to the expensive, high impact project only and 13% donating to both projects. Unlike behavioral adoption patterns in the field, the mode of donations goes to the high impact, high cost project.[16] One potential reason for this might be that the high impact project clearly has much more impact, even if underestimated on average. As pre-registered, I test whether self-serving beliefs are instrumental (i.e., serve the purpose of justifying donation decisions) by testing whether changes in incentives for accurate beliefs change donation patterns. In line with the finding that beliefs do not change across incentive structures, donation patterns also do not change significantly between *LoAB* and *HiAB* (Table 8 in the appendix).

I therefore take a more detailed look at whether beliefs are predictive of donation behavior, pooling over different incentive schemes (Table 2). Impact beliefs have a small but significant impact on the likelihood of donating to a specific project, in the expected direction.[17] A higher belief for the given project is positively correlated with the likelihood of donating, while a higher belief for the alternative project is negatively correlated with the likelihood of donating to the given project (although the latter is not robustly significant). This makes sense, given that projects' causes

[16] For example, in the environmental domain, only a small minority adopts a vegan diet (e.g., 2% for the UK population, YouGov (2022))

[17] The fact that beliefs conditional on true impact only have a small effect on donation likelihood might be an artifact of the experimental parameters, where differences in true impact are very large (compared to belief variance around the true impact level).

	1 if donated to low imp.		1 if donated to high imp.		total impact
	(1)	(2)	(3)	(4)	(5)
belief (true imp. = 8)	0.10***	0.11***	−0.05*	−0.06*	−0.06
	(0.02)	(0.02)	(0.02)	(0.02)	(0.07)
belief (true imp. = 32)	−0.03	−0.03	0.09***	0.09***	0.25***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)
altruism		0.17		0.29***	0.89***
		(0.09)		(0.08)	(0.22)
WTP 32 doses		0.06***		0.13***	0.43***
		(0.02)		(0.02)	(0.04)
Constant	0.10	−2.90	−3.04	−10.43***	−12.98*
	(2.73)	(1.95)	(1.78)	(2.37)	(5.66)
Session FE	Yes	Yes	Yes	Yes	Yes
Demographic controls	No	Yes	No	Yes	Yes
# subj.	607	607	607	607	607

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: RE-Logit regressions analyzing the role of beliefs for donations. The dependent variable in the first two models is a dummy with value 1 when donated to cheap low impact project (model 1 and 2), expensive high impact project (model 3 and 4). Model 5 is a linear RE-regression with total impact (vitamin A doses donated) as dependent variable with clustering robust SEs in parentheses. Beliefs are transformed into the implied belief of number of vitamin A doses. Willingness to pay for 32 doses is the switching point on the multiple price list (between 0 and \$40). For subjects, who did not switch, willingness to pay was coded as \$42 (the next step on the scale).

are perfectly substitutable. When looking at the total impact generated by donations (model 5), only beliefs for the high impact project significantly predict total impact. Finally, both the survey measure of general altruism (Falk et al. 2016) and the incentivized measure of willingness to pay has predictive power for a donation of 32 vitamin A doses.

If motivated beliefs are instrumental for donations, it could be that these reflect different types of donors rather than the same donors for different projects. For any given project, some subjects might want to *under*estimate impact to justify *not* donating, while others might want *over*estimate impact to justify donating. If this was the case, the belief effects of these different motives might cancel out on average when not taking the donation decision into account. Then, higher incentives for accurate beliefs might lead some subjects to downward correct their belief, and some to upward correct their belief, based on their initial donation choice. I therefore look at the interaction effect of accuracy bonus and donation choice on impact beliefs for the high and low impact project (Table 11 in the appendix). Donors have substantially higher beliefs than non-donors, ($\Delta = 23.97$ for the high impact project, $p < 0.001$; $\Delta = 17.35$ for the low impact project, $p < 0.001$; t-test). This pattern is interesting, but should be interpreted with caution. While it is in line with a self-serving interpretation of signals of donors and non-donors, I cannot rule out reverse causality (subjects

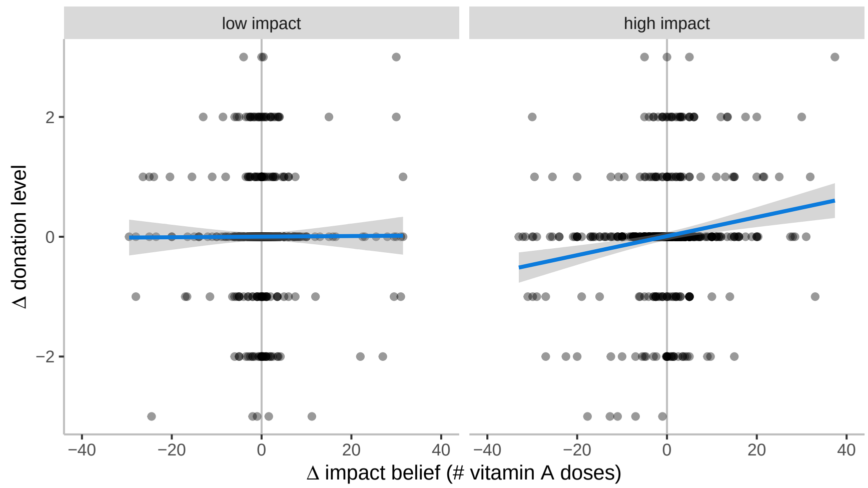


Figure 3: Correlation between changes in impact belief (x-axis) and donation levels (y-axis) (*HiAB* - *LoAB*). Donation levels are coded as 0 = no donation, 1 = donation to low impact project only, 2 = donation to high impact project only, 3 = donation to both projects.

with higher beliefs are more likely to donate). Additionally, subjects who donated in *LoAB* do not react significantly differently to a increase in incentives for accurate beliefs compared to non-donors (Table 11), providing additional evidence that motivated beliefs do not seem to play a role in my setting.

Finally, the within-subject nature of the change in incentives allows me to investigate the relationship between donations and beliefs further. I therefore look at how *changes* in beliefs within subject for a given project (i.e., the direction of updating as a response to higher incentives in *HiAB*) are correlated with *changes* in donation levels (Figure 3). Looking at donation levels, instead of donations to low and high impact projects separately, also accounts for income effects from changes in donations from one project to the other. It seems that for the low-impact projects, beliefs are already quite accurate, so increasing incentives does not change beliefs much even within subject (many observations are around a zero change in beliefs). For the high impact project, however, there is a positive correlation between the change in belief and the change in donation (those who correct their belief upwards donate more, those who correct downwards donate less). This suggests that changes in donation level are more driven by a change in high impact beliefs than low impact beliefs. Corresponding regression results can be found in the appendix (Table 12).

The small but significant correlation between changes in beliefs and changes in donations suggests that heterogeneity in beliefs may play a role in explaining donation patterns more broadly. This finding is also in line with findings from Metzger and Günther (2019), who show that subjects react differently to impact information, with some people donating less, some people donating more. My data provides suggestive evidence that this might be because they had different impact beliefs to start with and that impact information caused them to update in different directions.

3.3 Effect of *ExAnte* belief elicitation

Figure 4: Change in donation behavior in *ExAnte* compared to *ExPost* treatment

Subjects in the *ExPost* treatment discussed so far might still overestimate low impacts and underestimate high impacts as an ex post rationalization to justify their donation choices. To test whether scope for ex post rationalization does indeed affect beliefs, I compare *ExPost* impact beliefs to beliefs in the *ExAnte* treatment (between subject). In this treatment, subjects have to indicate their impact beliefs before making the donation decision, instead of after donating. This makes it more difficult for subjects to change their beliefs as a justification for their donation choices. However, *ExAnte* elicitation of beliefs does not affect impact beliefs on average. Low impact beliefs are on average 87.23 in *ExPost*, while they are only slightly lower in *ExAnte* (86.97; $p = 0.93$; two-sample t-test). Also, for the high impact project, differences are negligible and not significant (*ExPost*: 284.49; *ExAnte*: 286.55; $p = 0.62$; two-sample t-test). The pre-registered difference in differences is also descriptively very small and not statistically significant (Table 9 in the appendix).

While *ExAnte* elicitation of impact beliefs does not change beliefs on average, the change in procedure might still affect donation choices. Eliciting beliefs prior to donation decisions also forces subjects to think about impact before making their donation choice. Additionally, having one's own beliefs written down when making the donation choice may simplify comparison between and integration of cost and impact information across projects. For example, Toma and Bell (2022) find that adding decision aids such as presenting projects side by side or adding an impact calculator changed subjects' funding decisions and made them more sensitive to differences in impact. I therefore test whether there are differences in donation patterns between treatments. A chi-squared test confirms that distributions are indeed different (Figure 4, p = 0.0048). Looking closer at the donation data reveals that this is only driven by an increase in donations to the cheap, low impact project (light green and shaded areas, Figure 4). For this, I segregate the data, separating into two variables: donation frequency to the low impact project, and to the high impact project. While donations to the cheap project increase by 11.79 percentage points ($p < 0.001$), donations to the expensive project remain at a comparable level (Δ=-2.10 percentage points, p = 0.6; tests of proportions).

Finally, I look at how this plays out in the aggregate. The overall effect on average total donation amounts is a small, but insignificant increase (*ExPost*: \$9.15, vs. *ExAnte* \$9.29, p = 0.81; two-sample t-test). This can be reconciled with the previous findings by noting that the extensive margin (i.e.,whether people donate at all) remains almost unchanged (cf. the bottom group in Figure 4). Therefore, small increases in donations to either the cheap project only (decreasing the average total amount) or both projects jointly (increasing the average total amount) cancel each other out on average. One way to interpret this finding is by noting that in the *ExAnte* treatment, subjects see their impact beliefs for both projects when making their donation decision. This might make it easier to think about impact (or more difficult to ignore impact) and hence increase donations on the margin that is relatively cheap. In the next section, I take a closer look at how donation patterns are affected by making subjects think about impact first across different cost and impact combinations.

3.4 Robustness check: The *NoChoice* treatment

To conclusively rule out motivated beliefs as a driver of biased beliefs, I run a *NoChoice* treatment.[18] In this treatment, subjects see the same donation projects (i.e., prices and signals), but are only asked to indicate their impact belief. In the instructions, I make clear that subjects will *not* have to donate to any of these projects. Removing the choice between different donation decisions should remove the incentive to form motivated beliefs, as these beliefs are not tied to a consumption-utility affecting choice any more. If biased beliefs in the other treatments are indeed driven by bounded rationality instead of motivated cognition, as conjectured so far, I should not see a difference in beliefs between *ExPost* and *NoChoice*.

The graph below shows that already descriptively beliefs are very similar across all three treatments, for both incentive levels for accurate beliefs. Beliefs are on average 87.23 for the low impact project,

[18]This treatment was not pre-registered, and conducted in December 2022 as a robustness check of the results in the main experiment. Similar to the *ExAnte* treatment, I collected 296 observations on Prolific.

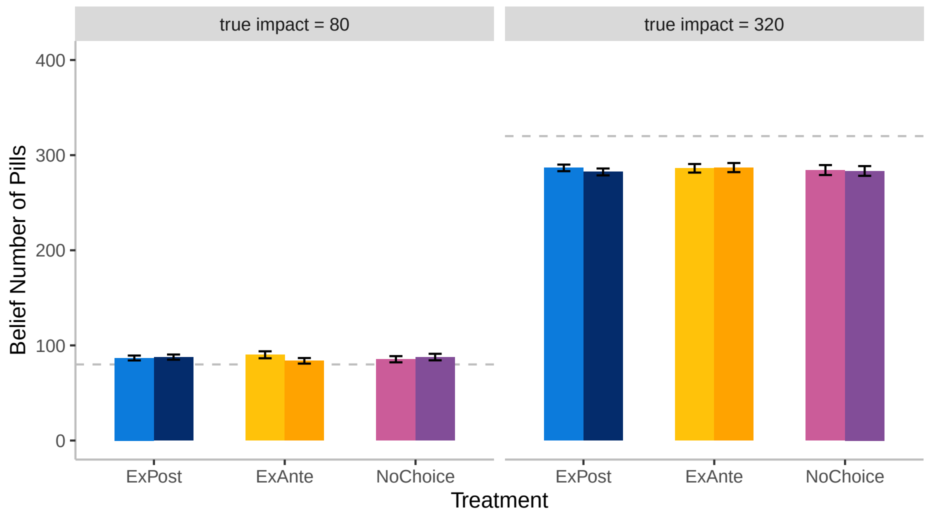


Figure 5: Mean impact beliefs across treatments. Lighter colors represent data from *LoAB*, darker colors represent data from *HiAB*. Whiskers depict standard errors.

and 284.49 for the high impact project. Two t-tests for the respective impact levels confirm that beliefs in the *NoChoice* treatment are not significantly different from beliefs in the *ExPost* treatment (p = 0.84 for low impact project, p = 0.89 for high impact project). The regression reported in the appendix also confirms this results when controlling for additional covariates (Table 10). These results corroborate the earlier finding that motivated cognition seems to play a limited role for impact belief formation in my setting.

4 Effect of different trade-offs

4.1 Experimental design

In the previous section, I focused on the case where subjects face a choice set of a cheap, low impact project and an expensive, high impact project. In the experiment, however, I systematically vary price and impact combinations to rule out that subjects could directly infer impact from prices (see Table 3). This variation allows me to run three additional exploratory analyses. First, they allow me to isolate the effect of prices on (motivated) belief formation. For example, I can compare the impact belief of project B, between parameter combination 1 and parameter combination 5, to see whether a high price leads to an underestimation of impact. Second, it allows me to investigate the role of the choice context (in particular the trade-off at hand) in the formation of motivated beliefs in more

	Project A			Project B		
price	# doses	efficiency	price	# doses	efficiency	channel
4	8	2	16	32	2	cost-impact trade-off (1)
4	8	2	16	**<u>8</u>**	0.5	cost trade-off (2)
4	**<u>32</u>**	8	16	32	2	cost trade-off (3)
4	8	2	**<u>4</u>**	32	8	cost level (4)
<u>16</u>	8	0.5	16	32	2	cost level (5)

Table 3: Experimental parameters. Bold-faced, underlined values represent changes compared to first row. Efficiency is given by impact per cost. In the main parameter combination of interest (1), I keep efficiency between the two projects constant. Row (1) represents a cost-impact trade-off, row (2) and (3) a cost trade-off, and row (4) and (5) aim to test effect of the cost level

detail. I hypothesize that the role for motivated beliefs in justifying donation choices is strongest when there is a clear impact-cost trade-off which makes subjects close to indifferent between the two projects. Tilting beliefs in the hypothesized direction (overestimation of cheap projects' impact and underestimation of expensive projects' impact) simplifies the trade-off as it makes the projects seem very different in terms of perceived efficiency, i.e., impact per cost. A simple trade-off between costs (without differences in impact) might also suffice to generate these hypothesized belief patterns. If this was the case, beliefs should be equally biased in parameter combinations where projects have the same impact level but differ in terms of costs (row 2 and 3). In particular this would imply that cheaper projects' impact should be overestimated and more expensive projects' impact should be underestimated, regardless of their true impact. Finally, the absolute cost *level* could be enough to create these directional belief changes, regardless of the monetary trade-offs between the options at hand. In such a case, one would expect to see overestimation when both projects are cheap (row 4), and underestimation of both projects' impact when both are expensive (row 5). Third, these structured changes in cost and impact combinations allow me to gain insights into how sensitive demand for donations is to changes in price and impact level. Specifically, they allow me to estimate cost and impact elasticities as well as cross-elasticities at different points along the demand curve for donations (e.g., price elasticities at different levels of impact).

4.2 Results

I start by comparing beliefs across the different parameter combinations. Beliefs are, however, robust to these changes in the choice context, and are on average not significantly different from each other across all parameter combinations (see Table 13 in the appendix). This corroborates the robustness of beliefs to changes in incentives for accuracy and suggests that motivated cognition does not play a role in determining beliefs in my setting.

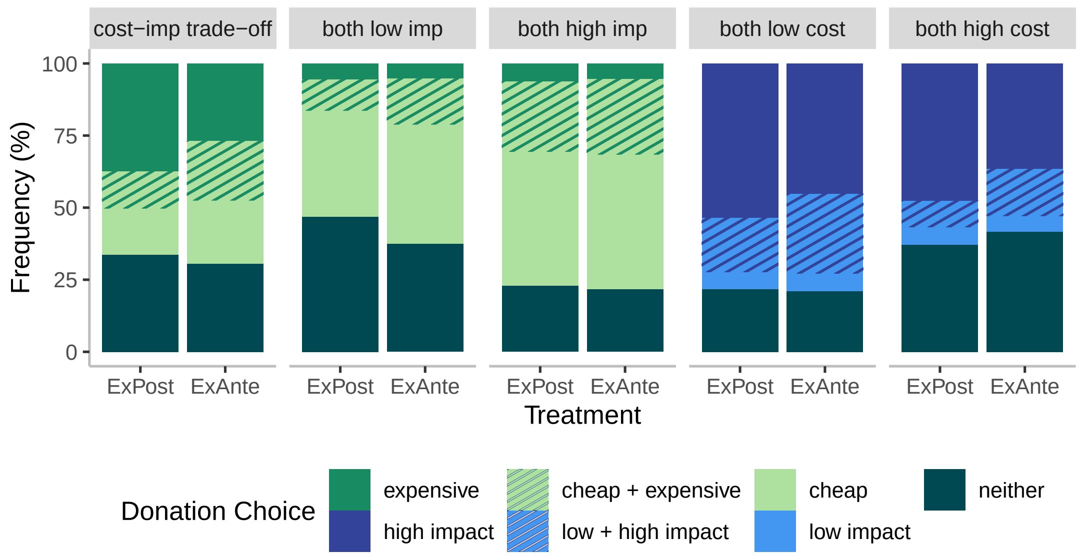


Figure 6: Donation patterns by treatment across the different parameter combinations in Table 3. Data is pooled across the different incentives for accurate beliefs.

I therefore move on to analyze how donation patterns react to the changes in incentives. I start by looking descriptively at donation patterns across the different parameter combinations (Figure **??**). First, it becomes clear that changes in trade-offs indeed change donation patterns in meaningful ways. Only a small fraction of subjects donate to options which are dominated by another alternative in the choice set (for utility functions that are weakly increasing in impact and money). In particular, on average 5% donate to the expensive project only, when the same impact level can be reached by a cheaper donation. Similarly, 6% donate to only the low impact project when a four times larger impact was available at the same cost. More generally, donations respond in a meaningful way to changes in incentives. Fewer people donate when donations are more costly (comparing panel 4 and panel 5), and there are more donations when large impact is relatively cheaper (i.e., in panel 3 and 4, where either both projects have high impact or both projects are cheap). Thus, subjects' insensitivity to changes in incentives for accurate beliefs is not merely driven by more general inattention to incentives.

4.2.1 Estimating demand elasticities

I design the parameter combinations in a structured way such that only one characteristic (price or impact) at a time would change, compared to the benchmark which involves a trade-off between impact and costs (parameter combination 1). These structured changes in project impacts and costs allow me to separately estimate demand elasticities (and cross-elasticities) with respect to changes in impact and cost. For example, comparing the frequency of donations to project B across parameter combinations (1) and (2) allows me to estimate impact elasticity for high price levels, since the only parameter that changes is the impact of the high price project B (cf. Table 3). As the elasticity may change along the demand curve, I can additionally estimate these elasticities separately for different impact and cost levels. That is, for example, by comparing donation frequency to project A between parameter combinations (1) and (3), I estimate impact elasticity also at low price levels. Similarly, I separately estimate price elasticity for the high and low impact project.

For the estimation, I run linear RE-regressions accounting for unobservable differences between subjects and controlling for the *HiAB* treatment. To be specific, I estimate the following RE-regression models exploiting the within-subject treatment variation in parameter combinations:

$$D_{ikt} = \beta_{0i} + \beta_1 \text{VOI}_{kt} + \beta_2 HiAB_t + u_i + e_{it}$$

where D_{ikt} is a dummy variable, taking value 1 when subject i donated to project k in period t. The variable of interest (VOI) varies with the different models. For elasticity estimates with respect to own characteristics, this variable was either $Price_k$ or $Impact_k$. For cross-elasticities, the regression specification used the variation in the VOI for the second donation project in the choice set, $Price_{-k}$, or $Impact_{-k}$. $HiAB_t$ is a dummy variable taking value 1, when the subject received a high accuracy bonus for accurate beliefs. Demand elasticity can then be derived from the regression coefficients, by multiplying the coefficient of interest β_1 (which captures how much demand changes with respect to changes in the VOI) with the average level of the VOI, divided by the average donation level to project k across the entire panel (to achieve a percentage change value for the elasticity estimate).[19] Results are robust to adding additional controls. To obtain confidence intervals for these elasticity estimates, I use non-parametric bootstrapping. Specifically, I re-sample the subject-level paired observations with replacement and estimate the models on this bootstrapped sample. The .025 and .975 percentiles of these 10,000 estimates yield the bounds of the 95% confidence intervals.

Figure 7 shows the estimated elasticities by treatment with bootstrapped 95%-confidence intervals (using non-parametric bootstrap). Several interesting results emerge from this analysis. First, when looking at own price elasticities (top left panel), it becomes clear that demand for donations is generally price inelastic (i.e., all estimates are below 1). That is, for a 1 percent increase in price, donations drop by less than 1 percent. Subjects react more to an increase in price of the low impact (-0.51 in *ExPost* and -0.53 in *ExAnte*) compared to the same price increase for high impact. High impact projects' price elasticities are estimated to be -0.30 in *ExPost* and -0.35 in *ExAnte*.[20]

[19] I estimate arc elasticities, that is taking the midpoint between the two prices (Allen and Lerner 1934). Unlike point elasticities, arc elasticity estimates only require the knowledge of demand behavior at two points (prices) and no assumptions on the functional form of the entire demand curve.

[20] These point estimates are at the lower end of what is typically obtained in the literature using field data (Adena,

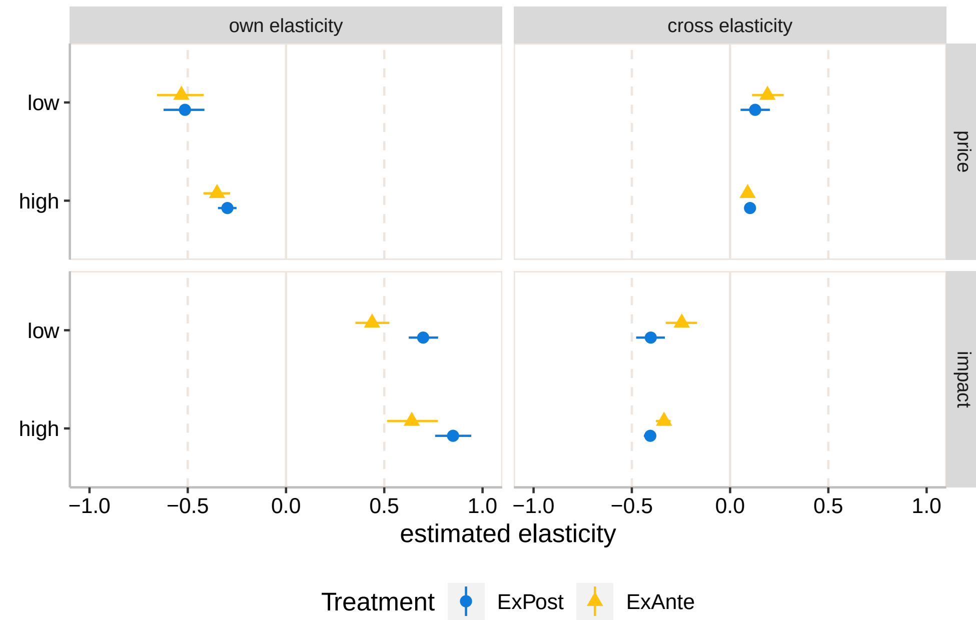


Figure 7: Elasticity estimates from linear RE-regressions with *HiAB* controls. The top boxes display these price elasticities for different impact levels, the bottom boxes display impact elasticities for different price levels. Errorbars display non-parametrically bootstrapped 95%-confidence intervals.

One explanation for this could be that lower impact projects are less attractive to donate to, so that differences in prices matter more, at least for the relatively stark differences in prices in my experiment.

Interestingly, estimated impact elasticities are larger than price elasticities, implying that subjects react stronger to changes in impact compared to proportional changes in prices (bottom left panel). Specifically, impact elasticities are estimated to be 0.85 in *ExPost* and 0.64 in *ExAnte*. In a similar vein to the observation of changes in price elasticity along the demand curve, impact elasticities are lower for low cost projects than for high cost projects (*ExPost*: 0.70; *ExAnte*: 0.44). Finally, I estimate cross-elasticities, that is how much demand for donating to one project changes as a response to a change in price or impact of the other project. As one would expect, cross-price and cross-impact elasticities are substantially smaller, with point estimates ranging between 0.09 and 0.19 for prices, and -0.41 and -0.25 for impact (right panel).

Hakimov, and Huck 2020). One potential explanation for this could be that subjects treat money in the experiment as a wind-fall gain, making them more likely to donate (and hence less sensitive to changes in prices).

4.2.2 Effect of *ExAnte* elicitation of impact beliefs on donations

I now move on to analyze how the order of belief elicitation affects donations across the different trade-offs. Comparing the elasticity estimates from the previous section, it becomes clear that own impact elasticities are on average 0.24 lower in *ExAnte* than in *ExPost*, while price elasticities are in a comparable range across treatments. At first sight, this seems at odds with results from Toma and Bell (2022), who show that giving participants tools that simplify impact comparison across projects *increases* impact elasticity. Taking a closer look at donation patterns in the *ExAnte* treatment reveals, that subjects are more likely to donate to both projects aggregating over different trade-offs ($p < 0.001$, test of proportions). This holds true for all parameter combinations but the one where both projects have high impact (Column 3, Table 4). For this parameter combination, donations to both projects are already close to their maximum level in *ExPost* (given subjects' preferences and the parameter combinations), with 24% donating to both projects (compared to 26% in *ExAnte*), suggesting that the non-significant effect for this parameter combination might be the result of a ceiling effect. These results also hold when applying a Bonferroni-correction for multiple hypothesis testing. Making subjects think about impact first, by eliciting their impact beliefs *ExAnte* before their donation choices, increases the likelihood to donate to both projects by 8% (Column 6 (all), Table 4).

	parameter combination (#)					
	(1)	(2)	(3)	(4)	(5)	overall
1 if *ExAnte*	0.08^{**}	0.05^{*}	0.02	0.09^{**}	0.07^{***}	0.08^{**}
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Constant	-0.08^{**}	-0.05^{*}	-0.02	0.16	-0.07^{***}	-0.04
	(0.02)	(0.02)	(0.03)	(0.18)	(0.02)	(0.04)
Bonferroni adjusted p-values	0.006	0.072	1.000	0.006	0.005	
# obs.	1810	1810	1810	1810	1810	9050
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
# subj.	905	905	905	905	905	905

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Linear RE-Regressions on likelihood of donating to both projects for all trade-offs (parameter combinations). Clustering-robust SE in parentheses. Model 6 looks at the overall effect for the full sample, controlling for all parameter combinations and their interaction effects with the *ExAnte* treatment. P-values that adjust for multiple hypothesis testing are reported at the bottom of the table (Bonferroni correction).

Eliciting beliefs before donation choices does not only make ex post rationalization more difficult (and hence affects beliefs directly). Subjects see their impact beliefs for the two projects next to the costs of donating to these projects. The fact that this change in choice environment has an effect on donations – making subjects more likely to donate to both projects – suggests that difficulties in integrating cost and impact information may be important in explaining how people choose between different other-regarding behaviors. This finding is in line with results from Toma and Bell (2022), who show in an experiment that cognitive uncertainty about how to integrate different dimensions of impact information can explain low sensitivity to differences in impact in the funding decisions of policy makers. They show that adding decision aids such as impact calculators increases impact sensitivity, corroborating the conjecture that complexity in comparing and aggregating impact information plays a decisive role in under-sensitivity to impact.

5 Discussion and Concluding Remarks

I conducted an online experiment to test the role of motivated impact beliefs in explaining the adoption of low and high impact other-regarding behaviors. Subjects in the experiment can donate to different projects which vary in impact and costs. However, while costs are always perfectly known, impact has to be estimated from a noisy signal using an attention task.

In line with what has been documented in the field, I provide evidence of a biased belief pattern. Subjects (slightly) overestimate low impacts on average and underestimate high impacts, with 22% of subjects displaying exactly this pattern. However, these beliefs are robust to changes in the decision environment. Neither changes in monetary incentives for accurate beliefs, nor changes in intrinsic incentives for motivate beliefs (by changes in costs of donation and trade-offs or by removing donation choices altogether) change beliefs significantly. Neither does eliciting beliefs prior to donation choices. This is somewhat unexpected, given previous results in the literature and theoretical predictions, which would postulate that changing trade-offs between material and belief-based utility would change the scope for motivated beliefs. One potential reason for this null result could be that impact levels are too different from one another in the present setting, making it too easy to recognize which project has high and which low impact. This could make it more difficult for subjects to convincingly make themselves believe that projects are comparable in terms of impact, and thus limit the scope for motivated beliefs. However, some recent papers also find that beliefs do not respond to changes in monetary incentives, or the order of elicitation (Engelmann et al. 2022; Gangadharan, Grossman, and Xue 2021a), or that impact beliefs are generally not motivated (Pace and Weele 2020), suggesting that further research is warranted to understand better when and how beliefs respond to financial and utility-based incentives.

While eliciting beliefs before donation choices does not affect beliefs, it does change donation patterns, making subjects 8 percentage points more likely to maximize impact by donating to both projects. One potential mechanism that could explain this effect is the cognitive difficulty in integrating cost information and impact beliefs across different projects. Nudging subjects to think about impact before making their donation decision by making their own beliefs salient makes this integration easier and thus increases the likelihood of maximizing impact. This finding is in line with results showing that choice aids such as presenting different projects next to each other or impact calculators which reduce cognitive uncertainty about impact levels make subjects funding choices more impact maximizing (Toma and Bell 2022). Further research is warranted to test how robust this finding is and to better understand more broadly the role of numeracy and complexity in determining impact sensitivity of other-regarding behaviors.

The results of this paper suggest a potential role for choice architecture to make donation or volunteering decisions more impactful. Nudging people to think about impact, and thus simply relying on beliefs is however not recommendable as a policy intervention, as impact beliefs in the field are often biased, even if not self-servingly so (Ipsos 2021). Rather, the results of my paper highlight the importance of choice aids such as carbon footprint calculators, or charity evaluators as used for example by GiveWell, which make impact information easily comparable across different behaviors. An interesting extension of this project would thus be a test of the welfare effects on other-regarding behavior of such information provision in a given context, say environmental behavior. Given my finding, and the findings from the field (Ipsos 2021) that not only underestimation of impacts is present, but also overestimation, impact information might lead donors to update their impact beliefs downwards, thus leading to fewer donations (Metzger and Günther 2019; Rodemeier and Löschel 2020).

References

Adena, Maja, Rustamdjan Hakimov, and Steffen Huck. 2020. "Charitable Giving by the Poor: A Field Experiment in Kyrgyzstan." Working Paper. WZB Discussion Paper.

Ahumada, Beatriz, Yufei Chen, Neeraja Gupta, Kelly Hyde, Marissa Lepper, Will Mathews, Silveus Neil, et al. 2022. "Well Excuse Me! Replicating and Connecting Excuse-Seeking Behaviors."

Alekseev, Aleksandr, Gary Charness, and Uri Gneezy. 2017. "Experimental Methods: When and Why Contextual Instructions Are Important." *Journal of Economic Behavior & Organization* 134: 48–59.

Allen, Roy George Douglas, and Abba Ptachya Lerner. 1934. "The Concept of Arc Elasticity of Demand." *The Review of Economic Studies* 1 (3): 226–30.

Ambuehl, Sandro. 2017. "An Offer You Can't Refuse? Incentives Change How We Inform Ourselves and What We Believe." Working Paper No. 6296.

AmeriCorps, Office of Research and Evaluation. 2021. "Key Findings from the 2019 Current Population Survey: Civic Engagement and Volunteering Supplement." AmeriCorps, Office of Research; Evaluation.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal* 100 (401): 464–77.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *The Quarterly Journal of Economics* 117 (3): 871–915.

Blanken, Irene, Niels van de Ven, and Marcel Zeelenberg. 2015. "A Meta-Analytic Review of Moral Licensing." *Personality and Social Psychology Bulletin* 41 (4): 540–58.

Bosch-Rosa, Ciril, Daniel Gietl, and Frank Heinemann. 2021. "Risk-Taking Under Limited Liability: Quantifying the Role of Motivated Beliefs." Working Paper.

Brunnermeier, Markus K, and Jonathan A Parker. 2005. "Optimal Expectations." *American Economic Review* 95 (4): 1092–1118.

Burdea, Valeria, and Jonathan Woon. 2022. "Online Belief Elicitation Methods." *Journal of Economic Psychology* 90: 102496.

Carpenter, Jeffrey. 2021. "The Shape of Warm Glow: Field Experimental Evidence from a Fundraiser." *Journal of Economic Behavior & Organization* 191: 555–74.

Caviola, Lucius, Stefan Schubert, and Joshua D Greene. 2021. "The Psychology of (in) Effective Altruism." *Trends in Cognitive Sciences.*

Caviola, Lucius, Stefan Schubert, Elliot Teperman, David Moss, Spencer Greenberg, and Nadira Faber. 2020. "Donors Vastly Underestimate Differences in Charities' Effectiveness."

Chater, Nick, and George Loewenstein. 2022. "The i-Frame and the s-Frame: How Focusing on the Individual-Level Solutions Has Led Behavioral Public Policy Astray." Working Paper.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting Moral Wiggle Room:

Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33 (1): 67–80.

Danz, David, Lise Vesterlund, and Alistair J Wilson. 2020. "Belief Elicitation: Limiting Truth Telling with Information on Incentives." Working Paper. National Bureau of Economic Research.

Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman. 2015. "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism." *American Economic Review* 105 (11): 3416–42.

Diekmann, Andreas, and Peter Preisendörfer. 2003. "Green and Greenback: The Behavioral Effects of Environmental Attitudes in Low-Cost and High-Cost Situations." *Rationality and Society* 15 (4): 441–72.

Dubois, Ghislain, Benjamin Sovacool, Carlo Aall, Maria Nilsson, Carine Barbier, Alina Herrmann, Sébastien Bruyère, et al. 2019. "It Starts at Home? Climate Policies Targeting Household Consumption and Behavioral Decisions Are Key to Low-Carbon Futures." *Energy Research & Social Science* 52: 144–58.

Engelmann, Jan, Maël Lebreton, Peter Schwardmann, Nahuel A. Salem-Garcia, and Joel J. van der Weele. 2022. "Anticipatory Anxiety and Wishful Thinking." Working Paper.

Enke, Benjamin, and Thomas Graeber. 2021. "Cognitive Uncertainty." Working Paper.

Enke, Benjamin, Ricardo Rodriguez-Padilla, and Florian Zimmermann. 2022. "Moral Universalism and the Structure of Ideology." *The Review of Economic Studies* forthcoming.

Exley, Christine L. 2015. "Excusing Selfishness in Charitable Giving: The Role of Risk." *The Review of Economic Studies* 83 (2): 587–628.

———. 2020. "Using Charity Performance Metrics as an Excuse Not to Give." *Management Science* 66 (2): 553–63.

Exley, Christine L, and Judd B Kessler. 2019. "Motivated Errors." Working Paper. National Bureau of Economic Research.

Eyster, Erik, Shengwu Li, and Sarah Ridout. 2021. "A Theory of Ex Post Rationalization." *arXiv Preprint arXiv:2107.07491.*

Falk, Armin, Anke Becker, Thomas J Dohmen, David Huffman, and Uwe Sunde. 2016. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences."

Filiz-Ozbay, Emel, and Neslihan Uler. 2019. "Demand for Giving to Multiple Charities: An Experimental Study." *Journal of the European Economic Association* 17 (3): 725–53.

Gangadharan, Lata, Philip J Grossman, and Nina Xue. 2021a. "Identifying Self-Image Concerns from Motivated Beliefs: Does It Matter How and Whom You Ask?" *Monash University Department of Economics Working Paper.*

———. 2021b. "Identifying Self-Image Concerns from Motivated Beliefs: Does It Matter How and

Whom You Ask?" *Monash University Department of Economics Working Paper.*

Genç, Murat, Stephen Knowles, and Trudy Sullivan. 2021. "In Search of Effective Altruists." *Applied Economics* 53 (7): 805–19.

Ging-Jehli, Nadja R, Florian H Schneider, and Roberto A Weber. 2020. "On Self-Serving Strategic Beliefs." *Games and Economic Behavior* 122: 341–53.

Gino, Francesca, Michael I Norton, and Roberto A Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives* 30 (3): 189–212.

GiveWell. 2021. "Helen Keller International's Vitamin a Supplementation Program." https://www.givewell.org/charities/helen-keller-international [Accessed: 07. Dec 2021].

Gneezy, Uri, Alex Imas, and Kristóf Madarász. 2014. "Conscience Accounting: Emotion Dynamics and Social Behavior." *Management Science* 60 (11): 2645–58.

Hagmann, David, Emily H Ho, and George Loewenstein. 2019. "Nudging Out Support for a Carbon Tax." *Nature Climate Change* 9 (6): 484–89.

Heeb, Florian, Julian F Kölbel, Falko Paetzold, and Stefan Zeisberger. 2022. "Do Investors Care about Impact?" *The Review of Financial Studies* Forthcoming.

Hestermann, Nina, Yves Le Yaouanq, and Nicolas Treich. 2020. "An Economic Model of the Meat Paradox." *European Economic Review* 129: 103569.

Hope Consulting. 2012. "Money for Good: The US Market for Impact Investments and Charitable Gifts from Individual Donors and Investors." Hope Consulting.

Indiana University Lilly Family School of Philanthropy. 2021. "The Giving Environment: Understanding Pre-Pandemic Trends in Charitable Giving." https://philanthropy.iupui.edu/news-events/news-item/latest-data-shows-new-low-in-share-of-americans-who-donated-to-charity.html?id=363 [Accessed: 21.Sept 2022].

Ipsos. 2021. "Perils of Perception: Climate Change." https://www.ipsos.com/en/ipsos-perils-perception-climate-change [Accessed: 14. Sept. 2022].

Jhunjhunwala, Tanushree. 2021. "Searching to Avoid Regret in Charitable Giving." Working Paper.

Lades, Leonhard K., Kate Laffan, and Till O. Weber. 2021. "Do Economic Preferences Predict Pro-Environmental Behaviour?" *Ecological Economics* 183: 1–10.

Maki, Alexander, Amanda R Carrico, Kaitlin T Raimi, Heather Barnes Truelove, Brandon Araujo, and Kam Leung Yeung. 2019. "Meta-Analysis of Pro-Environmental Behaviour Spillover." *Nature Sustainability* 2 (4): 307–15.

Metzger, Laura, and Isabel Günther. 2019. "Making an Impact? The Relevance of Information on Aid Effectiveness for Charitable Giving. A Laboratory Experiment." *Journal of Development Economics* 136: 18–33.

Niehaus, Paul. 2020. "A Theory of Good Intentions." Working Paper. UC San Diego.

Null, Clair. 2011. "Warm Glow, Information, and Inefficient Charitable Giving." *Journal of Public*

Economics 95 (5-6): 455–65.

Osman, Moustafa. 2011. "In-Kind Donations: Who Benefits?" 49. https://odihpn.org/publication/in-kind-donations-who-benefits/#irrelevant-aid [Accessed: 20. Sept 2022]: Humanitarian Exchange Magazine.

Ottoni-Wilhelm, Mark, Lise Vesterlund, and Huan Xie. 2017. "Why Do People Give? Testing Pure and Impure Altruism." *American Economic Review* 107 (11): 3617–33.

Pace, Davide, and Joël van der Weele. 2020. "Curbing Carbon: An Experiment on Uncertainty and Information about CO2 Emissions." Working Paper. Tinbergen Institute.

Pew Research Center. 2019. "U.s. Public Views on Climate and Energy." https://www.pewresearch.org/science/wp-content/uploads/sites/16/2019/11/PS.11.25.19_climate-energy-FINAL.pdf [Accessed: 20. Sep 2022].

Rodemeier, Matthias, and Andreas Löschel. 2020. "The Welfare Effects of Persuasion and Taxation: Theory and Evidence from the Field." *ZEW-Centre for European Economic Research Discussion Paper*, no. 20-019.

Saccardo, Silvia, and Marta Serra-Garcia. 2022. "Enabling or Limiting Cognitive Flexibility? Evidence of Demand for Moral Commitment." *American Economic Review* Forthcoming.

Tiefenbeck, Verena, Thorsten Staake, Kurt Roth, and Olga Sachs. 2013. "For Better or for Worse? Empirical Evidence of Moral Licensing in a Behavioral Energy Conservation Campaign." *Energy Policy* 57: 160–71.

Toma, Mattie, and Elizabeth Bell. 2022. "Understanding and Improving Policymakers' Sensitivity to Program Impact." Working Paper.

WHO. 2011. "Guideline: Vitamin A Supplementation in Infants and Children 6–59 Months of Age." Geneva: World Health Organization.

Woodford, Michael. 2020. "Modeling Imprecision in Perception, Valuation, and Choice." *Annual Review of Economics* 12: 579–601.

Xiang, Yang, Thomas Graeber, Benjamin Enke, and Samuel J Gershman. 2021. "Confidence and Central Tendency in Perceptual Judgment." *Attention, Perception, & Psychophysics* 83 (7): 3024–34.

YouGov. 2022. "Dietary Choices of Brits (e.g. Vegeterian, Flexitarian, Meat-Eater Etc)?" https://yougov.co.uk/topics/lifestyle/trackers/dietery-choices-of-brits-eg-vegeterian-flexitarian-meat-eater-etc [Accessed: 12.Sept 2022].

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–61.

A Sample demographics main experiment

Below I report the balance test of the elicited demographic variables and time spent in the experiment (i.e., in the *ExPost* and *ExAnte* treatments). The treatments are balanced with respect to all demographic variables. Some variables were elicited on Likert-scales (reported in the table below).

Variable		Levels	*ExPost*	*ExAnte*	p-value
Age	1	17 or younger	4.360	4.453	0.303
	2	18 - 19			
	3	20 - 29			
	4	30 - 39			
	5	40 - 49			
	6	50 - 59			
	7	60 or older			
Education	1	< High School	2.946	2.977	0.659
Level	2	High School			
	3	Bachelor			
	4	Master			
	5	Doctorate			
	6	other			
Female	1	Female	0.479	0.500	0.545
	0	Male			
Household	1	< $10,000	6.505	6.819	0.197
income	2	$10,000 - $19,999			
	3	$20,000 - $29,999			
	4	$30,000 - $39,999			
	5	$40,000 - $49,999			
	6	$50,000 - $59,999			
	7	$60,000 - $69,999			
	8	$70,000 - $79,999			
	9	$80,000 - $89,999			
	10	$90,000 - $99,999			
	11	$100,000 - $149,999			
	12	> $150,000			
Time spent		(minutes)	20.901	19.159	0.001

Table 5: Balance table for sample demographics. Levels of Likert-scale variables in second and third column. The fourth and fifth column display variable means in *ExPost* and *ExAnte*, respectively. The last column displays p-values of two-sided t-tests for differences in means across treatments.

B Within subject variation in beliefs

Below, I plot the within subject variation (s.d.) in beliefs by treatment. Specifically, I calculate for each subject the standard deviation in beliefs across rounds by treatment and true impact level. Importantly, only very few subjects show zero variation (3 subjects, 0.25% of the sample). At the same time, only few subjects vary a lot in their impact beliefs across rounds, which would be indicative of not paying attention at all.

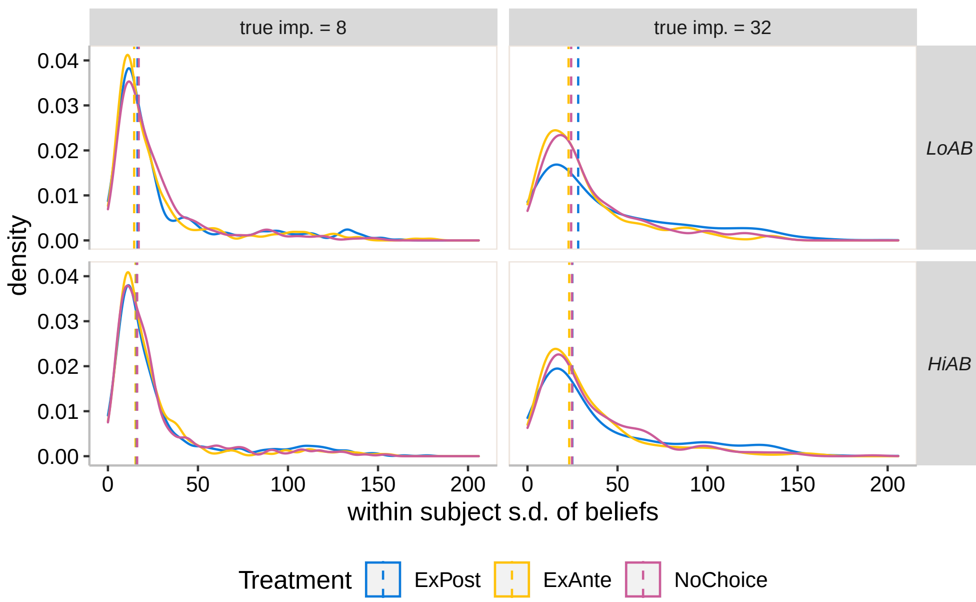


Figure 8: Density plots for within subject variation (s.d.) of impact beliefs by true impact and treatment. Dashed lines represent medians of the distributions.

C Analysis of cognitive uncertainty

I report the results of analyzing the effect of cognitive uncertainty (Enke and Graeber 2021) on belief formation. Cognitive certainty was elicited using a 20-step slider that returned implied confidence intervals which were narrower for higher values of certainty. Figure 9 shows densities of cognitive uncertainty in the all treatments. I continue the analysis on the next pages, testing whether cognitive uncertainty has an effect on belief formation. As neither cognitive certainty nor beliefs are differently distributed across the three treatments (p = 0.59 for difference in cognitive certainty between *ExPost* and *ExAnte*, p = 0.16 for difference between *ExPost* and *NoChoice*; Kolmogorov-Smirnov test), I continue the analysis pooling across all treatments.

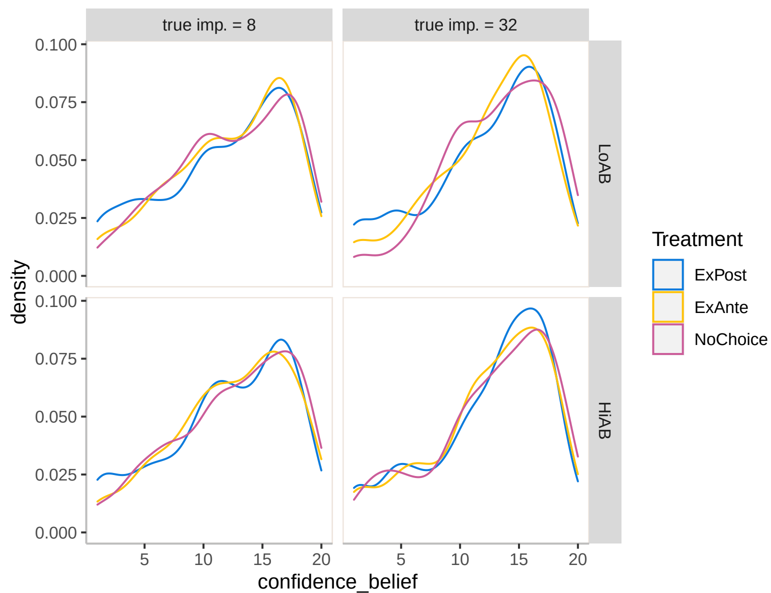


Figure 9: Densities of reported certainty in own belief in all treatments.

I test whether higher incentives for accuracy increase self-reported certainty in beliefs pooling across all three treatments. The regression results suggest that higher incentives for accuracy, weakly increase self-reported certainty of high impact beliefs. For low impact beliefs, there is no effect of higher incentives on perceived certainty of one's own estimate.

	low impact project		high impact project	
	(1)	(2)	(3)	(4)
1 if *HiAB*	0.21	0.21	0.26*	0.27*
	(0.14)	(0.14)	(0.13)	(0.13)
Constant	15.65***	16.59***	13.62***	14.18***
	(1.24)	(1.67)	(1.59)	(1.90)
# obs.	1810	1808	1810	1808
Session FE	Yes	Yes	Yes	Yes
# subj.	905	904	905	904

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 6: Linear RE-regressions with cognitive certainty as dependent variable. Cognitive certainty was elicited using a 20-step slider that returned implied confidence intervals which were narrower for higher values of certainty. Clustering-robust standard errors in parentheses.

In the next step, I look at whether cognitive certainty is correlated with absolute belief error pooling across all three treatments. The results in Table 7 indicate that self-evaluated certainty of belief accuracy is correlated with a lower error in belief, at least for the high impact project. However, higher incentives for accurate beliefs do not change this significantly (insignificant main and interaction effect). This suggests that the incentive effect on beliefs is too small to have an additional effect through cognitive uncertainty.

	low impact project			high impact project		
	(1)	(2)	(3)	(4)	(5)	(6)
Cogn. certainty	−0.46	−0.49	−0.41	−2.16***	−2.13***	−2.15***
	(0.27)	(0.32)	(0.32)	(0.40)	(0.45)	(0.46)
Cogn. certainty x 1 if *HiAB*		0.07	0.11		−0.10	−0.09
		(0.36)	(0.36)		(0.50)	(0.50)
1 if *HiAB*		−0.83	−1.39		4.76	4.66
		(4.87)	(4.92)		(7.08)	(7.08)
Constant	30.96***	31.34***	11.43	89.70**	87.55**	83.45**
	(4.30)	(4.86)	(8.21)	(28.70)	(29.04)	(30.77)
# obs.	1803	1803	1801	1806	1806	1804
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	No	No	Yes	No	No	Yes
# subj.	905	905	904	905	905	904

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7: Linear RE-regressions with absolute belief error as dependent variable (|belief-true impact|). Cognitive certainty was elicited using a 20-step slider that returned implied confidence intervals which were narrower for higher values of certainty. Clustering-robust standard errors in parentheses.

D Pre-registered analysis: Effect of *LoAB* on donation

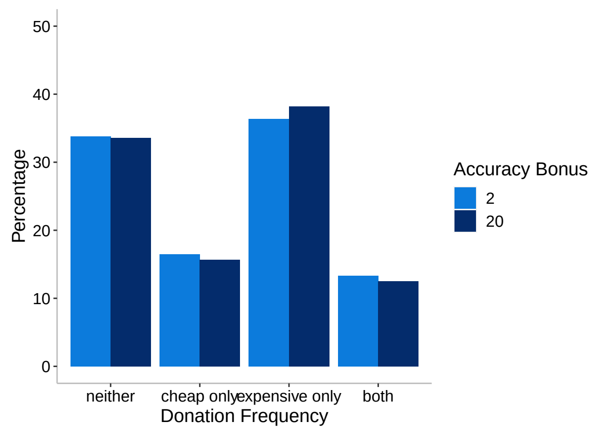


Figure 10: Donation distributions in *ExPost* for the two accuracy incentives

I pre-registered to test whether donations are different in the treatment that allows for motivated beliefs (*LoAB*) compared to the treatment where it is more difficult to maintain motivated beliefs because incentives for accuracy are higher (*HiAB*). I first run a chi-squared test comparing overall donation distributions that are depicted in Figure 10 ($p = 0.91$). As pre-registered, I run two separate RE logit regressions for donations to the low and the high impact project with a dummy that takes value 1 when there are low incentives for accurate beliefs. As beliefs do not change across treatments, low incentives for accurate beliefs do also not affect donations significantly (Table 8).

	1 if donated to low imp.		1 if donated to high imp.	
	(1)	(2)	(3)	(4)
1 if *LoAB*	0.20	0.12	-0.11	-0.26
	(0.20)	(0.38)	(0.19)	(0.73)
impact belief		0.01^{***}		0.01^{***}
		(0.00)		(0.00)
impact belief x *LoAB*		0.00		0.00
		(0.00)		(0.00)
Constant	-0.10	-3.36	-0.82	-9.76^{***}
	(2.60)	(1.78)	(1.97)	(2.48)
Session FE	Yes	Yes	Yes	Yes
Demographic controls	No	No	No	Yes
# subj.	607	607	607	607

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8: RE-Logit regressions analyzing the role of beliefs for donations.

E Between subject treatment effects on beliefs

I pre-registered to test for the difference in differences in beliefs across treatments to see whether ex post rationalization strengthens motivated beliefs. In Table 9, I run this test in a linear panel regression framework, using clustering robust standard errors. The results show that there is no main effect of *ExAnte* elicitation on beliefs. The difference in differences between *HiAB* and *LoAB* and *ExPost* and *ExAnte* is also not significant.

	low cost/impact		high cost/impact	
	(1)	(2)	(3)	(4)
ExAnte	3.63	2.81	−0.32	−0.51
	(4.48)	(4.45)	(5.69)	(5.72)
HiAB x *ExAnte*	−7.15	−7.19	4.95	4.99
	(4.66)	(4.66)	(5.68)	(5.68)
Constant	55.70***	32.40***	274.94***	284.98***
	(3.88)	(7.88)	(35.76)	(39.07)
# obs.	1803	1801	1806	1804
Session FE	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes
# subj.	905	904	905	904

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 9: Linear RE-regressions including session FE with belief as dependent variable testing for a difference in differences in beliefs using treatment dummies. Clustering-robust standard errors in parentheses.

I run the same regressions to test for an effect of the *NoChoice* treatment on impact beliefs.

	low cost/impact		high cost/impact	
	(1)	(2)	(3)	(4)
NoChoice	−1.22	−0.27	−2.57	−2.63
	(4.13)	(4.13)	(6.27)	(6.27)
HiAB x *NoChoice*	1.42	1.18	3.51	2.75
	(4.49)	(4.51)	(5.82)	(5.80)
Constant	86.69***	54.85***	286.61***	299.36***
	(2.56)	(8.47)	(3.46)	(11.86)
# obs.	1800	1794	1798	1792
Session FE	No	No	No	No
Demographic Controls	No	Yes	No	Yes
# subj.	903	900	903	900

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 10: Linear RE-regressions with belief as dependent variable testing for a difference in differences in beliefs using treatment dummies. Clustering-robust standard errors in parentheses. Session FE are omitted due to collinearity with the treatment.

F Relationship between beliefs and donations

In the following, I look at the interaction effect between incentives for accurate beliefs and donations. In particular, subjects who donate in the condition with low incentives for accurate beliefs may overestimate impact, and consequentially downward adjust their belief when facing higher incentives for accurate beliefs. At the same time, subjects who do not donate may underestimate impact and hence upward adjust their belief when facing higher incentives for accurate beliefs ($HiAB$). These two effects might cancel out on average leading to the null-result on the pre-registered test on differences in beliefs (Figure 11).

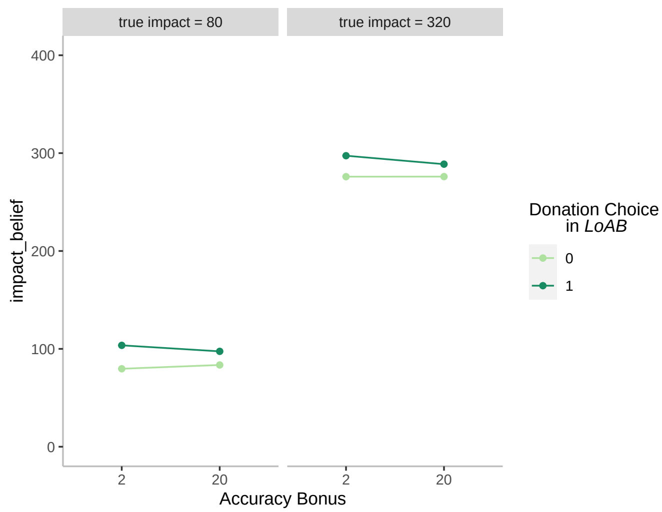


Figure 11: Interaction effect between donation choices and beliefs

In Table 11, I test for this interaction. Using donation decisions in the *LoAB* treatment, I test whether donors update their belief differently than non-donors. While the results go in the predicted direction, they do not reach significance, suggesting that there is only a main effect of donors having a higher belief than non-donors.

	low cost/impact		high cost/impact	
	(1)	(2)	(3)	(4)
HiAB	3.89	3.91	0.22	0.14
	(3.09)	(3.10)	(5.36)	(5.37)
1 if donated in *LoAB*	23.81^{***}	22.60^{***}	21.57^{**}	22.09^{**}
	(6.64)	(6.58)	(6.84)	(6.89)
HiAB x 1 if donated in *LoAB*	-9.85	-9.84	-8.85	-8.75
	(6.80)	(6.80)	(7.29)	(7.30)
Constant	74.95^{***}	40.71	297.81^{***}	332.56^{***}
	(22.71)	(24.10)	(10.98)	(16.94)
# obs.	1209	1207	1210	1208
Session FE	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes
# subj.	607	606	607	606

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 11: Linear RE-regressions including session FE with belief as dependent variable looking at effect of donation choices in the *LoAB* treatment on beliefs. Clustering-robust standard errors in parentheses.

In Table 12, I test whether there is within subject correlation between the direction of updating beliefs and the direction of changes in donation level, when controlling for individual differences. That is, I look at by how much subjects' donation level (in the sense of not donating, donating to low impact only, to high impact only, or to both projects) changes in response to a change in impact beliefs. The results suggest that the relationship is indeed significant, although the effect size is small.

	Δ donation level				
	(1)	(2)	(3)	(4)	(5)
Δ low impact belief	0.00	0.00			0.01
	(0.01)	(0.01)			(0.01)
Δ high impact belief			0.02^{***}	0.02^{***}	0.02^{***}
			(0.00)	(0.00)	(0.00)
Constant	-0.34	-0.25	-0.39	-0.32	-0.34
	(0.39)	(0.42)	(0.36)	(0.39)	(0.41)
Session FE	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes	Yes
# obs.	602	601	603	602	597

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 12: OLS-regressions with difference in donation levels as dependent variable. Clustering-robust standard errors in parentheses. Δ belief: Change in impact belief (high accuracy bonus - low accuracy bonus) in terms of number of doses donated.

G No context effect on beliefs

The table below shows average beliefs in all treatments (columns) for the different payoff combinations (rows). Each row reflects impact beliefs for a specific parameter combination (price, impact). As can be seen from the table, beliefs are descriptively very similar across the different treatments and parameter combinations and not significantly different using two-sided t-tests (even without a correction for multiple hypothesis testing), suggesting that neither prices themselves, nor potential contrasting effects induced by different prices in the choice set play a role in belief formation in my experiment.

Table 13: Average impact beliefs by treatment and project type

		ExAnte		*ExPost*		*NoChoice*	
		LoAB	HiAB	LoAB	HiAB	LoAB	HiAB
cost-imp trade-off							
(\$4, 80)	avg	90.14	83.79	86.76	87.70	85.47	87.75
	sd	63.46	50.76	63.02	66.88	55.79	58.89
(\$16, 320)	avg	286.16	286.93	286.61	282.36	284.34	283.38
	sd	77.80	82.80	85.27	89.08	89.88	88.04
both low impact							
(\$4, 80)	avg	84.35	86.00	80.93	84.04	86.77	82.01
	sd	55.77	57.30	59.21	64.00	57.52	50.72
(\$16, 80)	avg	83.72	85.23	84.40	85.57	88.93	80.99
	sd	50.67	53.47	59.81	62.39	63.22	46.93
both high impact							
(\$4, 320)	avg	284.02	293.49	287.97	286.58	289.23	290.04
	sd	80.90	78.50	87.55	88.22	84.81	81.35
(\$16, 320)	avg	285.16	290.54	291.67	290.71	289.15	291.81
	sd	79.77	78.55	85.46	81.64	84.70	80.98
both low cost							
(\$4, 80)	avg	86.91	83.64	89.83	88.83	87.11	85.87
	sd	57.14	52.95	70.24	66.23	57.30	53.99
(\$4, 320)	avg	289.11	290.83	278.64	282.94	282.11	283.28
	sd	76.52	78.91	93.68	91.77	87.80	85.07
both high cost							
(\$16, 80)	avg	90.72	84.47	91.20	87.37	90.28	88.57
	sd	64.18	58.86	69.66	64.47	58.04	56.12
(\$16, 320)	avg	284.22	290.85	285.63	287.36	285.08	285.52
	sd	79.68	76.33	88.50	87.06	85.94	88.56

H The decision screen

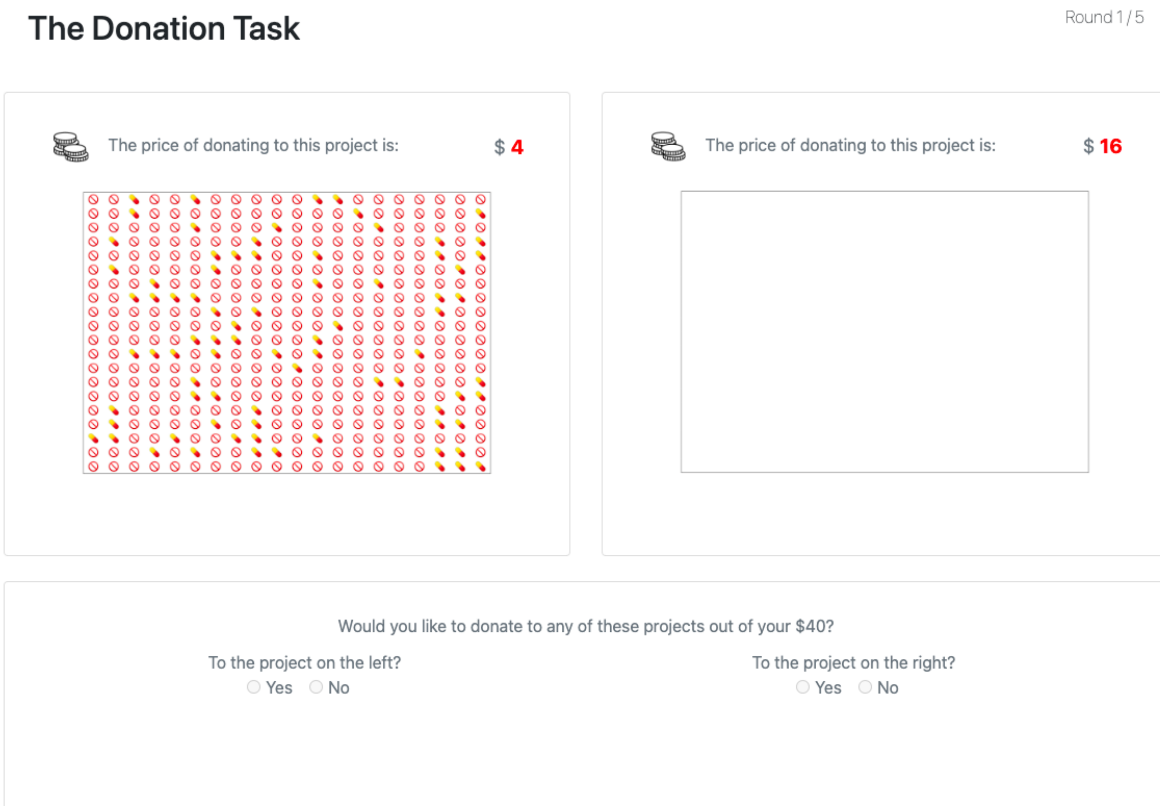


Figure 12: A screenshot of the decision screen

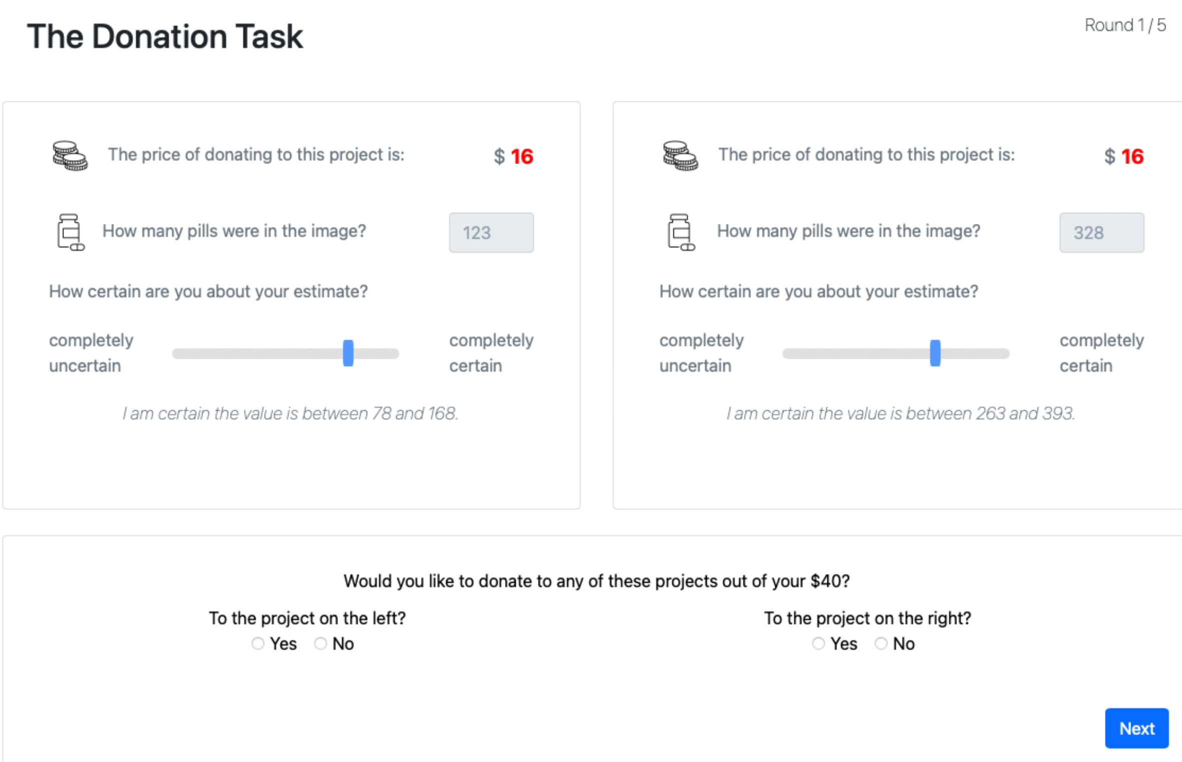


Figure 13: A screenshot of the decision screen after filling in beliefs in *ExAnte*

I Instructions

Welcome to this study!

Maastricht University

Welcome | Your Data | Informed Consent

Welcome

Welcome to this experiment conducted by Frauke Stehr, a researcher at Maastricht University. In the experiment you will be able to earn money based on the choices you make. **You will earn a fixed participation fee of €3.50 for completing the study.** You may additionally receive a bonus payment, depending on your choices in the experiment.

In case you have questions about this research, you can reach the researcher at

Frauke Stehr
Tongersestraat 53
6211 LM Maastricht
The Netherlands
Or via e-mail: f.stehr@maastrichtuniversity.nl

The experiment will take an average of 22 minutes. It is important to note that we apply **a strict "no-deception" rule**, meaning that **everything we write is true and implemented exactly as described**. The experiment was ethically approved by the Ethical Review Committee of Inner-City Faculties of Maastricht University.

Welcome to this study!

Maastricht University

Welcome | Your Data | Informed Consent

Your Data

What information do we collect and why?

Upon your consent, we collect responses to the experiment you provide. No individual can be personally identified from the collection of this data. The data is anonymous, with the exception of your Prolific ID, which is required to administer payments. The researcher does not have access to any other personal data and therefore cannot link this ID with any other personal data. After the payment has been processed, we will separate the Prolific IDs from the dataset and thereby anonymize it. The Prolific IDs will be stored separately and safely on a Maastricht University server. After this anonymization, it will not be possible to identify participants based on their responses.

The data collected in this experiment will be used for academic research in behavioural economics. We will publish the results in academic papers. These papers will contain only anonymous data presented as tables and figures (e.g., percentages and averages).

Cookies policy

We do not use cookies. We do store some technical information (e.g., your screen resolution) automatically to be able to analyse technical problems.

What is the legal basis for holding these data?

The lawful basis for processing this information is Article 6(1.e) of the General Data Protection Regulation (GDPR): processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.

Your data will not be used for other purposes. Only fully anonymized data will be shared for the sole purpose of replicating the analysis if requested by scientific journals. You have the right to request access to your data and/or deletion of your Prolific ID from our data by sending an email to f.stehr@maastrichtuniversity.nl.

How do we store the data?

The study uses the app oTree, which is hosted on the servers of Heroku, a cloud platform based in San Francisco (USA) and owned by Salesforce. To send emails to you the study uses the services of Academic Prolific. These companies declare to be compliant with the GDPR. No personal data other than your Prolific ID will be shared with these companies. Your Prolific ID is stored encrypted in the Heroku cloud database and cannot be read by anyone outside the research team. All raw data collected in the experiments is stored securely on servers at Maastricht University. Maastricht University stores all data for at least 10 years. After that, the data is destroyed or transferred to other media for longer storage if needed.

The faculty's Privacy Officer is Eric Soemers. For questions or complaints about the privacy legislation of this research you can contact him at SBE Administration Office – IT and Facilities, via h.soemers@maastrichtuniversity.nl.

Welcome to this study!

Maastricht University

Welcome | Your Data | Informed Consent

Your consent to participate

I hereby give permission for my data to be used for scientific purposes. I have had enough time to decide whether I want to participate in the experiment. I know that participation is voluntary, and I know that I can decide to withdraw from the experiment at any time. I do not have to justify such a decision to withdraw. If I withdraw I will forfeit any payments.

I understand that the data will be stored anonymously and thus can only be published anonymously. I give permission for the researcher to use my anonymized responses in subsequent experiments.

I understand that for people with photosensitivity, the characteristics of some of the images in the experiment may cause dizziness or other symptoms.

Before you begin, please turn off your phone/e-mails/music so that you can concentrate on this experiment. Thank you!

Participate in this study

Instructions

The structure of the experiment

- In this experiment, you can donate vitamin A supplements to children in need.
- The experiment is composed of two parts.
- Each part consists of multiple rounds of a task that is almost the same between the two parts.
- Your choices in these tasks **may affect your final earnings**.
- We randomly select 1 in 20 participants at the end of the experiment.
- If you are among them, we **randomly select one of your decisions** in the experiment to be the **decision-that-counts for your bonus payment**.

Throughout the rest of the instructions, we use yellow boxes to **explain how your decisions affect your bonus payment**.

- On the next pages, we describe the task.
- Pay close attention to these instructions, we will ask **comprehension questions** at the end of each section.
- You will **not be admitted to the experiment and forfeit any payments**, if you **answer incorrectly more than two times**.

Next

Instructions

Before we start, we need to ensure that you understand the instructions.
You may view the instructions again by going back to the previous page (click 'back' below).

If you are randomly selected for a bonus payment, your bonus payment depends on your answers in the experiment:

- True
- False

At most how many times are you allowed to answer wrongly to be admitted to the experiment?

Back

Check my answers and continue

Instructions - The Donation Task

What is your task in the experiment?

- You can **donate to different charity projects** which provide vitamin A supplements to children in need.
- The projects **differ in price and number of vitamin A supplements** financed.
- However, you do not perfectly know how many vitamin A doses your donation finances.
- Instead, you have to **estimate this number**.

⚠ Your donation has real consequences!

For each vitamin A dose you donate, we will provide the equivalent funding to a real charity for buying and administering vitamin A doses.

Why vitamin A supplements?

- Over 200,000 children's deaths can be attributed to vitamin A deficiency each year.
- The **World Health Organization recommends** that all preschool-aged children in areas where vitamin A deficiency is a public health problem receive vitamin A supplements two to three times per year.
- In this experiment, you can donate to children in need through Helen Keller International's Vitamin A Supplementation Program.[1]

Next

[1] You can read more about the charity and its scientific evaluation here.

Instructions - The Donation Task

How can you donate?

In each round,

- you receive $40, which you can spend on donations.
- you will see two projects.
- you can **donate to one, both, or neither** of the projects.
- your bonus payment may be determined by your donation(s).

If your donation decision in a given round is selected as the decision-that-counts, your bonus payment will determined as follows:

- If you donated to at least one of the two projects,
  - you **receive $40 minus the sum of prices of the projects to which you donated**, and
  - we will **donate vitamin A doses** to each of the projects to which you donated.
- If you decided not to donate, you will keep the entire **$40**.

How do donation projects differ?

- Each project is described by:
  - an image depicting the **number of vitamin A doses** this project finances (which you have to estimate),
  - **the price** of donating to this project
- The price can be **either $4 or $16**.

How many vitamin A doses your donation finances **depends only on the true number of pills** in the image and **not on your estimate**!

Back

Next

Instructions - The Donation Task

Before we start, we need to ensure that you understand the instructions.
Remember, if you answer **more than two questions wrongly**, you will be excluded from the experiment. You may view the instructions again by going back to the previous page (click 'back' below).

How many vitamin A doses your donation finances depends on...

- ○ your estimate of the number of pills in an image.
- ○ the true number of pills in an image.
- ○ whether your estimate is correct.

In what way does your donation decision have real consequences (if it is selected as a decision-that-counts)?

- ○ The money you donate will go to a charity of your choice.
- ○ Your choice has no real-world consequences.
- ○ The money you donate will go to another participant in this study.
- ○ If you donate, you fund real vitamin A doses to be administered by Helen Keller International.

Back

Check my answers and continue

Instructions - The Donation Task

How do you estimate the number of vitamin A doses?

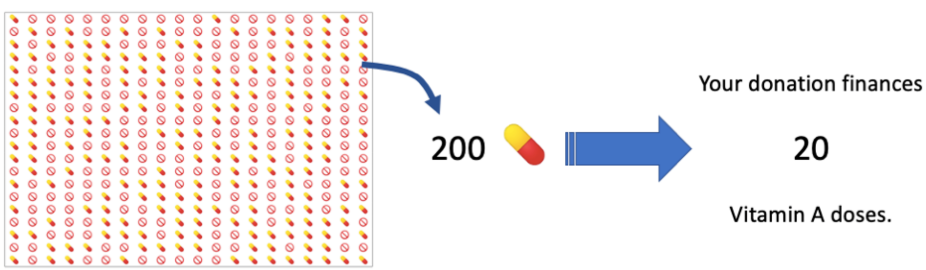
- For each project, you will see an image like the one below.
- You will see each image for **7 seconds**.
- Then, we ask for **your estimate** for the number of pills and how certain you are about your estimate.
- There are **between 0 and 400 pills** in an image.
- The number of pills 💊 indicates the number of vitamin A doses:
  for each 10 pills depicted on an image, your donation finances **1 vitamin A dose**.

If one of your estimates is selected as a decision-that-counts,

- you receive **$2** if your estimate is **at most 10 pills away from the true number** of pills.
- you receive **no bonus** if your estimate is **more than 10 pills away** from the true number.

⚠ We will never give feedback on how close your estimate was to the true number of pills.

Below you see an example of an image which contains 200 💊. In this project, your donation would then finance 20 vitamin A doses.



Next

Instructions - The Donation Task

Before we start, we need to ensure that you understand the instructions.
Remember, if you answer **more than two questions wrongly**, you will be excluded from the experiment. You may view the instructions again by going back to the previous page (click 'back' below).

How many pills are there in a given image?

- ○ at least 10 pills
- ○ Between 0 and 400 pills
- ○ at most 200 pills
- ○ This cannot be known

To make sure you have read the instructions, we ask you to answer 'apple' in the field below.

Suppose your estimate is 6 pills away from the true number of pills. What will be your bonus payment (in $) if this estimate is selected as decision-that-counts?

Back

Check my answers and continue with a trial round