Eliciting Moral Preferences: Theory and Experiment

Roland Bénabou Armin Falk Luca Henkel Jean Tirole

May 29, 2022

Abstract

We study the extent to which a person's moral preferences can be inferred from their choices, and how behaviors that appear deontologically motivated should be interpreted. Comparing direct elicitation (DE) and multiple-price list (MPL) mechanisms, we characterize how image motives inflate the extent of prosocial behavior. The resulting signalling bias is shown to depend on the interaction between elicitation method and visibility level: it is greater under DE for low reputation concerns, and greater under MPL for high ones. We test the model's predictions in an experiment with life-saving donations and find the key crossing effect predicted by the theory.

JEL codes: C91, D01, D62, D64, D78.

Keywords: Moral behavior, deontology, utilitarianism, consequentialism, social image, self-image, norms, preference elicitation, multiple price list, experiments.

Affiliations: Bénabou: Princeton University, NBER, CEPR, CIFAR, briq, IZA, BREAD, and THRED. Falk: Institute on Behavior and Inequality (briq) and University of Bonn. Henkel: University of Bonn. Tirole: Toulouse School of Economics (TSE) and Institute for Advanced Study in Toulouse (IAST), University of Toulouse Capitole.

Acknowledgments: We are grateful to Ingela Alger, Jean-François Bonnefon, Gary Charness, Franz Ostrizek, Pëllumb Reshidi and Marie-Claire Villeval for valuable comments. Ana Luisa Dutra, Juliette Fournier, Pierre-Luc Vautrey, Ben S. Young, Youpeng Zhang and Egshiglen Batbayar provided superb research assistance.
Funding: Bénabou gratefully acknowledges financial support from the Canadian Institute for Advanced Study, Tirole and Falk from the European Research Council (European Community's Seventh Framework Program Grant Agreement no. 249429 and no. 340950, as well as European Union's Horizon 2020 research and innovation programme, Grant Agreement no. 669217).
Ethics approval: The study was approved by the ethical committees of the University of Bonn (no. 2019-01), Toulouse School of Economics and Princeton University (no. 11818).

"In the kingdom of ends everything has either a price or a dignity. What has a price can be replaced by something else as its equivalent; what on the other hand is above all price and therefore admits of no equivalent has a dignity" (Kant, 1785).

1 Introduction

What can be inferred about people's moral preferences from observing their choices, and conversely how can decision environments be structured to maximize public-goods contributions or welfare? Relatedly, how should one interpret behaviors that seem deontologically rather than consequentially motivated, such as refusing tradeoffs that involve explicit harm to others, or assigning an infinite price to certain "sacred values" such as life, freedom, integrity of the human body, and dignity?

Empirically, behavior appears to reflect a variable mix of utilitarian, deontological and image-seeking motives that remains imperfectly understood. We examine these issues with a combined model and experiment in which agents incur a cost to do good, or forfeit a "bribe" for causing harm, under a variety of choice conditions. Three main lessons emerge.

First, as soon as image concerns are present, aggregating information from multiple choices decreases prosocial behavior, through what we term a *discouragement effect* that makes pooling more difficult. Conversely, the uncertainty inherent in ex-ante pledges increases prosocial behavior through two effects that reduce the effective cost of signaling, which we term *cheap-talk* and *cheap-act* effects. Second, all three effects are at work when comparing standard preference measurement schemes such as direct elicitation (*DE*) and multiple-price lists (*MPL*), leading to systematically different results. The same applies to public-goods provision and charitable contribution mechanisms that also feature multiple choices and uncertainty. Third, this behavioral divergence not only varies in magnitude, but even in sign, as the visibility or salience of choices increases. The model's most distinctive prediction is thus a "crossing" pattern. When image concerns are low (but positive) the discouragement effect dominates, so that for any given price, *DE* will generate more prosocial behavior than *MPL*. When image concerns are high the *cheap-act* effect dominates, resulting in *MPL* now generating more prosocial behavior – a full reversal. In particular, image-minded consequentialists will display Kantian-like price insensitivity much more readily under *MPL* than under *DE*.

Mechanisms and intuitions. We start from a basic model of prosocial behavior in the presence of image concerns, which we extend to study agents' decisions across a range of settings that differ in their *choice sets*, *material consequences* (the effective cost to oneself or others) and *reputational stakes.* Such variations occur naturally in a cross-section of individuals, or when observing someone's behavior in multiple situations. They also underlie the paradigm

of controlled experiments, leading us to center the analysis around the two most commonly used preference elicitation schemes, namely *DE* and *MPL*.

To introduce the model and the mechanisms at work, we use a simple example. Alex and Bob encounter their friend Chris, who works for an organization that recruits volunteers to tutor disadvantaged children in public schools (PS). The work is the same in all of the schools, every Friday afternoon, as are the students. The only difference is the commute: PS10 only requires a 10 minute round trip, PS30 half-an-hour, PS40 requires 40', and PS50 demands 50'. In the population of potential tutors, everyone has the same opportunity cost of time, but there are two equally likely types of social preferences, represented respectively by Alex, who would be willing to commute up to $v_H = 39$ minutes to tutor, and Bob, with a maximum willingness of $v_L = 27$ minutes.

Chris is in charge of recruiting for PS30, and she asks her two friends whether she can sign them up. Alex immediately accepts. Bob's first thought is to decline, but for a net cost of 3' per week he can look just as caring as Alex, and will thus accept if this image gain of $(v_H - v_L)/2$ is worth even just 6' (say) of actual travel[1]. If instead of Chris, the two friends had encountered Casey, who is recruiting for PS40, Alex would again have accepted but Bob would have declined: for a net cost of 1', Alex reaps an image gain of $v_H - v_L$ worth 12' of travel, which Bob cannot match given his net cost of 13'. In this "separate questions" setting, which corresponds to direct elicitation in experiments, even minor reputational concerns generate both pooling (PS30) and separating (PS40) that increase prosocial behavior.

Suppose now that Chris and Casey are working as a pair, so that Alex and Bob meet them at the same time. Casey first asks about signing up for PS40, then Chris about PS30: a format intermediate between direct elicitation and a standard multiple-price list. The fact that Alex agrees to PS40 while Bob declines is now sufficient to distinguish them, making it futile for Bob to volunteer for PS30, since there are no intermediate types (equivalently, he only cares about looking as good as Alex). By aggregating information from both choices, pooling has been eliminated and prosocial contributions reduced. This is the *discouragement effect,* which underlies our result that DE dominates MPL for positive but relatively low degrees of image concerns.

To understand the *cheap-act effect,* imagine that instead of sending out separate recruiters for each school, the organization has them ask everyone they approach which among the four schools they would be willing to tutor at. Each volunteer will then be randomly allocated, with equal chances, to one of those on their "acceptable" list. As before, the richness of options offered generates a discouragement effect for PS30. Working in the other direction, however, is the fact that listing any school up to PS40 now has an expected cost of only $80/3 < 27$', so Bob is willing to match Alex at that level. In fact, both friends will end

[1]The reputational stakes in this example are $(v_H - v_L)/2 = 6$ under pooling and $(v_H - v_L) = 12$ for separation; we normalize here the relative price of image to 1, for convenience.

up volunteering *unconditionally*. The expected cost of accepting all schools, including PS50, is only $140/4 = 32.5$', so clearly Alex will do it, but Bob will follow suit, as his net cost of pooling is only 5.5', while achieving that image is again worth 6'. By offering high-minded commitments that one will sometimes be called upon to fully honor, and sometimes "with a discount", the *cheap-act* effect increases prosocial contributions at *all* levels, relative to separate questions. Moreover, the stronger are reputational concerns, the stronger this effect. Indeed, if closer friends or a larger group will learn of one's choices, the acceptance threshold naturally increases; the expected cost below any cutoff grows more slowly than the cutoff itself, hence reflecting a complementarity.

The standard experimental implementation of multiple-price lists brings in another related *"cheap-talk"* effect, namely the possibility of not being called upon to contribute at all. Suppose that for every volunteer, the organization will draw at random (e.g., receive a phone call from) any of the four schools. If it is in the volunteer's "acceptable" list on the form that they filled out, they are assigned to that school; if not, they are not solicited, and moved to the bottom of the pile. The probability that a pledge may end up having been pure cheap talk makes everyone more willing to sign up ex-ante, again increasing contributions at every level.[2] As reputational concerns intensify, however, acceptance cutoffs rise, so this effect progressively vanishes. In particular, anyone listing all four schools as acceptable will be called upon with certainty, just as when acceding to a single-school request. While important in general (at intermediate values), this *cheap-talk* effect thus plays little role when image concerns are sufficiently important.[3] The cheap-act effect, in contrast, operates most powerfully in this case, and is the reason why contributions are then higher under *MPL* than under *DE*.

Experiment. The paper's second contribution is to test these novel predictions about how elicitation schemes and image concerns *interact*, using an experiment in which nearly 700 participants face significant moral decisions. Each subject's choice is to either: (i) direct a 350€ donation to a designated charity in India that will use the money to treat five tuberculosis patients, resulting statistically in the expected saving of one human life; or (ii) take money for themselves, where the amount is either a fixed 100€ under *DE*, or determined by the subjects' cutoff on an *MPL* where prices range from 0 to 200€. These two elicitation conditions are crossed with low and high moral-image treatments, allowing us to measure the distributions of behaviors for all four relevant outcomes. Comparing the fractions of subjects choosing the "saving a life" contribution over taking 100€, we find a clear and statistically significant reversal between *DE* and *MPL* as image concerns go from weak to strong, just as

[2]It operates like random implementation (which increases contributions under both *DE* and *MPL*, essentially scaling up all reputational payoffs), but differs from it in having an *endogenous* probability of implementation.

[3]Grossman (2015) studies the effect of implementation probability in a binary-choice dictator game, finding it to be relatively weak.

predicted by the theory. In the *Low Image* treatments, the fraction opting to save a life is 48% under MPL, versus 59% under DE*;* while in the *High Image* condition it is 63% under DE versus 72% under *MPL*.

We also conduct a placebo experiment with 366 additional subjects, where we keep all relevant aspects of the decision environment, but choices are over a non-moral good. Image concerns are now absent, and, as expected, we find no significant difference between the two elicitation methods.

1.1 Related Literature

The paper connects several distinct lines of work. The first centers on signaling and identity concerns in the moral domain (Bénabou and Tirole, 2006; Ellingsen and Johannesson, 2008; Ariely et al., 2009; Bénabou and Tirole, 2011a,b; DellaVigna et al., 2012; Lacetera et al., 2012; Ashraf et al., 2014; Exley, 2016; Gino et al., 2016; Grossman and van der Weele, 2017; Bursztyn et al., 2018; Falk et al., 2018; Bénabou et al., 2020). The second is concerned with the impact of being pivotal, whether in reality or in a "deontologically imagined" situation à la Kant (Brekke et al., 2003; Roemer, 2010; Alger and Weibull, 2013; Falk and Szech, 2013; Elias et al., 2016; Ambuehl, 2017; Bartling et al., 2020; Falk et al., 2020). In analyzing the interactions between the two, the paper also relates to work on auctions with signaling, in which bidders seek to demonstrate goodness, wealth, or a strong aftermarket position (Goeree, 2003; Giovannoni and Makris, 2014; Bos and Pollrich, 2020; Bos and Truyts, 2022).

With respect to experimental methodology, both the model and the experiment contribute to the study of alternative elicitation mechanisms. There is a fair amount of research comparing how *DE*, *MPL* or random implementation (strategy method, Selten, 1967) affects behavior in one-shot, anonymous games such as dictator or public-goods (Brandts and Charness, 2011; Chen and Schonger, 2016a).[4] There is also a large body of research on elicitation methods for risk and time preferences (Charness et al., 2013; Cox et al., 2015; Cohen et al., 2020; Baillon et al., 2022). To our knowledge, no such study has explored reputationally sensitive decisions like those analyzed here.

Finally, the paper relates to the (long-standing) debate in moral philosophy between consequentialist and deontological ethics. The evidence on how people behave in practice is mixed: the literature on cooperation and voluntary contributions to public goods finds that choices are generally sensitive to the implied consequences (Kagel and Roth, 1995; Goeree

[4]Concerning DE with deterministic versus random implementation (an intermediate case relative to MPL), the overview by Charness et al. (2016) reports generally ambiguous effects. As the model will make clear, it is only in the presence of sufficient signaling concerns that probabilistic implementation will matter. In contrast, risk attitudes play no role in the effects that we identify, which directly affect expected returns.

et al., 2002); similarly, charitable giving decreases when the risk of having no impact rises (Brock et al., 2013), or when overhead increases (Gneezy et al., 2014). At the same time, there is evidence of "warm glow" altruism in which utility seems to be derived from the act as such, with donations being fairly insensitive to the level of contributions by others (e.g., Andreoni, 1989, 1990) or the impact on the recipient (DellaVigna et al., 2022). Experiments in which subjects choose between money and a charitable act under varying probabilities that their decision will be implemented also point to a mix of consequentialist and rule-based or expressive motives (Feddersen et al., 2009; Chen and Schonger, 2016b; Falk et al., 2020).

2 Model

Preferences. Agents are risk-neutral, with a two-period horizon, $t = 1, 2$. At date 1, an individual can engage in prosocial behavior $(a = 1)$ or act selfishly $(a = 0)$. Choosing $a = 1$ involves a personal cost $c > 0$ but generates a public good or externality $e \geq 0$. Agents differ in their intrinsic motivation to act morally: given e, it is either $v_H e$ (high type) or $v_L e$ (low type), with probabilities ρ and $1 - \rho$, $v_H > v_L \geq 0$, and average $\bar{v} = \rho v_H + (1 - \rho) v_L$.

Besides the externality, the second feature of action $a = 1$ tying it to the moral domain is that it can be reputationally valuable, conferring a social or self-image benefit at date 2. In the social context, the agent knows his type but the audience (peer group, firms, potential partners) does not. In the self-signaling context, he has an immediate, "intuitive" sense of his deep preferences at the moment of action – for instance, how much empathy or spite he experiences – but later on the intensity of that feeling is imperfectly accessible ("forgotten"), and only the deed itself, $a = 0$ or 1, can be reliably recalled to assess his own moral identity.

Under either interpretation, an agent of type $v = v_H, v_L$ has expected utility

$$(ve - c)\, a + \mu \hat{v}(a), \tag{1}$$

where $\hat{v}(a)$ is the expected type conditional on the action $a \in \{0, 1\}$ and the circumstances under which it took place (deterministic cost, random draw from a list, etc.), while μ is the strength of self or social-image concerns, common to all agents. This utility may be additively augmented by any externalities generated by others, but since that term is independent of the agent's action we omit it here. Note that these preferences are consequentialist: an agent's desire to behave prosocially trades off the externality he expects his actions to have, the personal costs involved, and the reputational consequences.[5]

[5]The model also allows for genuinely deontological agents, $v_H = +\infty$, but a key point is that they are not needed to generate "observationally deontological" behavior in MPL-like experiments.

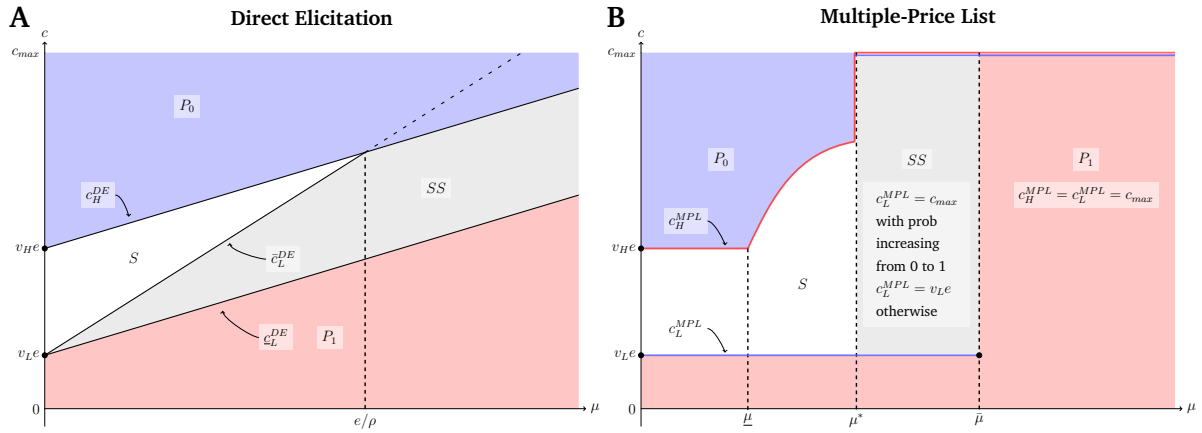

Figure 1: Equilibrium under Direct Elicitation (panel A) and Multiple-Price List (panel B). P_0: pooling at $a_H = a_L = 0$; S: separation, $a_H = 1, a_L = 0$; SS: semi-separation: $a_H = 1, a_L \in (0,1)$; P_1: pooling at $a_H = a_L = 1$.

Behavior. As common in signaling models, multiple equilibria may coexist: when

$$\max\{v_L e - c + \mu(v_H - v_L),\ v_H e - c + \mu(v_H - \bar{v})\} \leq 0 \leq v_H e - c + \mu(v_H - v_L),$$

there is both a pooling equilibrium at $a = 0$ and a separating one in which the v_H type contributes, with a mixed-strategy one in-between. In case of multiplicity (see Appendix A), we choose the equilibrium that is best for both types, namely the no-contribution pooling equilibrium. Indeed, separation yields lower payoffs for both, since $\mu v_L < \mu\bar{v}$ and $v_H e - c + \mu v_H \leq \mu\bar{v}$. This simple framework readily implies that an agent is more likely to act morally the higher the perceived externality e and/or his image concern μ.

3 Measuring Moral Preferences

We now expand the model to study what can be learned from people's (un)willingness to accept different tradeoffs between personal gain and harm to others. Two "canonical" choice situations are considered.

3.1 Direct Elicitation

Under *DE,* the individual faces a take-it-or-leave-it opportunity to incur a given cost (or forfeit a given prize) c to create an external benefit e. This decision can be made under varying levels of reputational concern μ, corresponding to the public visibility or private memorability of choices.

Equilibrium behavior will be characterized by three cost thresholds, strictly increasing

in μ and illustrated in Panel A of Figure 1 (for the case $\rho < 1/2$) :

$$v_H e - c_H^{DE}(\mu) + \mu\,(v_H - \bar{v}) \equiv 0, \tag{2}$$

$$v_L e - \bar{c}_L^{DE}(\mu) + \mu\,(v_H - v_L) \equiv 0, \tag{3}$$

$$v_L e - \underline{c}_L^{DE}(\mu) + \mu\,(\bar{v} - v_L) \equiv 0. \tag{4}$$

Denoting $a_H^{DE}(c,\mu)$ and $a_L^{DE}(c,\mu)$, or a_H and a_L for short, the two types' strategies (probabilities of choosing $a = 1$), one easily shows the following results.

Proposition 1. *The outcome of direct elicitation is as follows:*

1. *For low costs,* $c < \min\{\underline{c}_L^{DE}, c_H^{DE}\}$, *everyone behaves morally,* $a_H = a_L = 1$.
2. *For intermediate costs,* $c \in (\underline{c}_L^{DE}, c_H^{DE})$, *the high type behaves morally* $(a_H = 1)$, *but the low type's probability* $a_L(c)$ *of doing so decreases with* c, *and then equals* 0 *for* $c \geq \min\{\bar{c}_L^{DE}, c_H^{DE}\}$.
3. *For high costs,* $c \geq c_H^{DE}$, *both types behave immorally,* $a_H = a_L = 0$.

Relative to "pure" (intrinsic) moral preferences ve, decision thresholds are inflated due to reputational concerns; see (2)-(4). In particular, the range of costs $[\bar{c}_L^{DE}, c_H^{DE}]$ where full separation occurs shrinks with μ, becoming empty for $\mu > e/\rho$.

3.2 Multiple-Price List

Under *MPL,* the individual "names his price" by stating what maximum cost $c \in [0, c_{\max}]$ he is willing to incur for taking action $a = 1$, where $0 \leq v_L e < v_H e < c_{\max}$. Equivalently, c represents his willingness to accept a "bribe" to make the immoral choice, $a = 0$. This elicitation is made incentive-compatible by drawing some $\tilde{c} \in [0, c_{\max}]$ according to a pre-announced distribution $G(\tilde{c})$, and making him take the action at cost $\tilde{c}$ only when $\tilde{c} \leq c$. In experiments, G is typically uniform, but we allow any other case, including $c_{\max} = +\infty$. Let $L(c)$ denote the low type's net loss from selecting the cutoff c :

$$L(c) \equiv \int_{v_L e}^{c} (\tilde{c} - v_L e)\, dG(\tilde{c}),$$

and assume that $L(c_{\max}) < \infty$, a weak condition since it suffices that $E_G[\tilde{c}] < \infty$. We will say that a subject is *observationally deontological* if he turns down all prices on the proposed list (with distribution G): given the available data, he behaves as someone who would not act immorally "at any price."

We now solve for both types' willingness to accept (WTA) under the multiple-price list, denoted $c_H^{MPL}(\mu)$ and $c_L^{MPL}(\mu)$, or c_H^{MPL} and c_L^{MPL} for short. Note first that, absent reputation concerns ($\mu = 0$), MPL and DE are equivalent, and reveal the true preference: $c_H^{DE} = c_H^{MPL} = v_H e$, $\underline{c}_L^{DE} = \bar{c}_L^{DE} = c_L^{MPL} = v_L e$. For $\mu > 0$, comparing $L(c)$ to the reputational stakes $\mu(v_H - v_L)$ and $\mu(v_H - \bar{v})$ yields both types' equilibrium strategies, illustrated in Panel B of Figure 1 and characterized again by critical thresholds between separating, semi-separating and pooling regions:

$$\underline{\mu} \equiv \frac{L(v_H e)}{v_H - v_L} < \mu^* \equiv \frac{L(c_{\max})}{v_H - v_L} < \frac{L(c_{\max})}{\rho(v_H - v_L)} \equiv \bar{\mu}. \tag{5}$$

Proposition 2. *The outcome of the MPL mechanism is as follows:*

1. *When the (self) reputational concern μ is low, $\mu < \mu^*$, the high type's WTA for behaving immorally is $c_H^{MPL} = \max\{v_H e,\ L^{-1}(\mu(v_H - v_L))\}$, while the low type finds it too costly to pool and accepts $c_L^{MPL} = v_L e$.*

   Initially, for $\mu \leq \underline{\mu}$, separation is costless for the high type, then as μ rises he has to raise his reservation price to separate from the low type.

2. *When μ is intermediate, $\mu \in [\mu^*, \bar{\mu}]$, the high type can no longer separate and becomes observationally deontological, $c_H^{MPL} = c_{\max}$. The low type randomizes, with probability $a_L(\mu)$ increasing in μ, between that same "virtuousness" ($c_L^{MPL} = c_{\max}$) and revealing himself (accepting $c_L^{MPL} = v_L e$).*

3. *When $\mu > \bar{\mu}$, (self) image concerns are strong enough that both types' behavior is observationally deontological: $c_H^{MPL} = c_L^{MPL} = c_{\max}$.*

3.3 Comparison of DE vs. MPL

Under both elicitation schemes, image concerns naturally raise contributions, as seen in Figure 1. More novel and complex, however, are the following questions:

1. Is one elicitation scheme *more image-sensitive* than the other?

2. Which one yields *more expected contributions?*

Formally, at a given cost $c \in [0, c_{\max}]$, what fraction of people $\bar{a}^{DE}(c, \mu)$ accept forfeiting c to implement $a = 1$ under *DE,* versus what fraction $\bar{a}^{MPL}(c, \mu)$ state a willingness to pay of at least c under *MPL*? And how does $\bar{a}^{DE}(c, \mu) - \bar{a}^{MPL}(c, \mu)$ depend on μ?

While the answers generally depend on the specific value of c, the cases of sufficiently low and high image concerns yield clear predictions. We will denote as μ^{**} the solution to

$\underline{c}_L^{DE}(\mu) = c_{\max}$, or

$$\mu^{**} \equiv \frac{c_{\max} - v_L e}{\bar{v} - v_L} > \frac{L(c_{\max})}{\bar{v} - v_L} = \bar{\mu}. \tag{6}$$

Putting together the results of Propositions 1 and 2, we have:

Proposition 3. *For each type $\tau = H, L$,*

1. *For any $c \in [0, c_{\max}]$, $a_\tau^{DE}(c,\mu)$ and $a_\tau^{MPL}(c,\mu)$ coincide at $\mu = 0$, then both increase (weakly) as μ rises, reaching 1 for μ large enough.*
2. *For all $\mu \in (0, \underline{\mu})$, $a_\tau^{DE}(c,\mu) \geq a_\tau^{MPL}(c,\mu)$, with strict inequality for $c \in (v_L e, \bar{c}_L^{DE}(\mu))$ and $c \in (v_H e, c_H^{DE}(\mu))$, both nonempty.*
3. *For all $\mu \geq \bar{\mu}$, $a_\tau^{DE}(c,\mu) \leq a_\tau^{MPL}(c,\mu) = 1$, with strict inequality for $c \in (\underline{c}_L^{DE}(\mu), c_{\max})$, which is nonempty whenever $\mu \in (\bar{\mu}, \mu^{**})$.*
4. *The average behavior over types, $\bar{a}^m(c,\mu) \equiv \rho a_H^m(c,\mu) + (1-\rho) a_L^m(c,\mu)$, $m = DE, MPL$, inherits these same properties.*

The first result is standard, while the others stem from the interplay of three effects.

Weak image concerns: discouragement effect dominates. When $\mu > 0$ is low enough that separation under *MPL* is costless, we have $c_H^{MPL}(\mu) = v_H e < c_H^{DE}(\mu)$ and $c_L^{MPL}(\mu) = v_L e < \underline{c}_L^{DE}(\mu)$, hence the second result. Intuitively, *MPL* raises the cost to the low type of mimicking the high one, since to do so he must forego up to $v_H e$, and for low reputational gain such a discrete cost is not worth it. Under *DE*, in contrast, he pays only in proportion to the gain. This intuition is reflected in the fact that the lower boundary of the separating region is linear in Panel A of Figure 1, whereas it is initially flat in Panel B.

Strong image concerns: cheap-act effect dominates. At high values of μ, reputational concerns become paramount, and the cost of signaling is lower under *MPL* than under *DE,* since high values of c must only be paid with a probability less than 1: the effective cost of stating a cutoff c is only $E[\tilde{c}|\tilde{c} \leq c] < c$. It is even bounded by $L(c_{\max}) + v_L e < \infty$, which limits the extent to which the high type can separate, so that for $\mu > \bar{\mu}$ full pooling occurs: $c_H^{MPL} = c_L^{MPL} = c_{\max}$, so $a^{MPL}(c,\mu) = 1$, whereas $\bar{a}_L^{DE}(c,\mu) < 1$ as long as $\mu < \mu^{**}$. Most importantly:

Property 1. *For any distribution satisfying the monotone likelihood ratio property $(g/(1-G)$ increasing), the "discount" $c - E[\tilde{c}|\tilde{c} \leq c]$ is increasing in c. Therefore, as μ rises and with it each type's cutoff, the* cheap-act *effect becomes stronger, which increases contributions relative to* DE.

Intermediate image concerns. Inside $(\underline{\mu}, \bar{\mu})$, a third *"cheap-talk"* effect comes into play. Under *MPL,* an agent who states a cutoff $c < c_{\max}$ has only a probability $G(c) < 1$ of being

called upon to actually "deliver": if $\tilde{c} > c$ is drawn, she neither incurs a cost *nor* generates the externality e. This makes it safer to state high cutoffs, thus adding to the *cheap-act* effect. The latter is not as strong in this range as for high values of μ, and conversely the *cheap-talk* effect weakens as μ rises, pushing $G(c^{MPL})$ closer to 1. The net balance of the three effects is generally ambiguous in this intermediate range, and consequently so is the sign of $a^{DE} - a^{MPL}$ and its variation with μ.

Testable Implications. Three main hypotheses emerge from the model. First, as usual, greater visibility increases contributions. Second, at low but positive levels of visibility, *DE* leads to more prosocial outcomes, as the *discouragement effect* dominates. Third, at high levels (but not so high as to push everyone to $a = 1$ under *DE*), this ordering reverses: *MPL* induces more moral decisions, due to the now dominating *cheap-act* effect.

The inequalities in Proposition 3 can be weak or strong, depending on the region of the parameter space. This is a standard feature of models with discrete types and action spaces, which typically disappears when there is sufficient heterogeneity to span all cases. Thus, integrating over c:

Corollary 1. *Let an individual or population be confronted with a full-support distribution of costs $G(c)$, one by one under* DE*, or via a single list under* MPL*. The resulting contributions $\bar{a}_G^{DE}(\mu)$ and $\bar{a}_G^{MPL}(\mu)$ vary as in Proposition 3, but now with strict inequality.*

Real subject populations will also display much more heterogeneity in values v than in our model, resulting in similar smoothing. For this reason, when confronting the model with data, we will tighten the predicted inequalities to be strict ones.[6]

4 Experimental Design

4.1 Saving a Life

We adopt the *Saving a Life* paradigm from Falk and Graeber (2020), in which subjects can either take money for themselves or implement a fixed, life-saving donation to a charity dedicated to the treatment of tuberculosis in India. According to the World Health Organization, tuberculosis is one of the ten leading causes of death worldwide, even though there are highly effective antibiotic treatments available with success rates well above 85%. Together with the Indian non-profit organization *Operation ASHA*, we calculated a specific monetary

[6]Our tests will vary $(\tau, \mu) \in \{DE,\ MPL\} \times \{\mu_L, \mu_H\}$ while maintaining the same realized cost c. By Proposition 3, the set of parameters such that $0 < \mu_L < \underline{\mu} < \bar{\mu} < \mu_H < \mu^{**}$ and $\bar{a}^{DE}(c, \mu_L) - \bar{a}^{MPL}(c, \mu_L) > 0 > \bar{a}^{DE}(c, \mu_H) - \bar{a}^{MPL}(c, \mu_H)$ is nonempty provided that $\underline{c}_L^{DE}(\bar{\mu}) < c_H^{DE}(\underline{\mu})$, which reduces to $L(c_{\max}) - (1-\rho)L(v_H e) < (v_H - v_L)\, e$. Sufficient conditions are easily found; with a uniform G, for example, $c_{\max}/e \in \left(v_H - \left(v_H^2 - v_L^2\right)^{1/2}, v_H + \left(v_H^2 - v_L^2\right)^{1/2}\right)$.

amount sufficient to identify, treat, and cure a number of patients such that – in expectation – one patient will be saved from death by tuberculosis due to the donation. Combining public information on the charity's operations with estimates from peer-reviewed epidemiological studies on tuberculosis mortality for the specific type of treatment and location considered (Kolappan et al., 2008; Straetemans et al., 2011; Tiemersma et al., 2011), we determined that level to be 350€: by allowing for the treatment of five patients, such a donation allows the (expected) saving of one human life.

This paradigm contrasts the option of saving a life (major positive externality e) by triggering a donation of 350€ versus that of taking money for oneself (opportunity cost c), inducing a clear tradeoff between morality and self-interest. This is reinforced by the donation being cost-effective: the amount is well above all monetary payments possible for the subjects themselves, as described later, and the money is directly used to treat patients, without any administrative or transaction cost.

4.2 Treatments

To test the model's predictions, we use a 2×2 between-subjects design, varying the elicitation method (DE vs. MPL) as well as the visibility and moral salience of choices (*Low Image* vs. *High Image*) at the payment stage.

Under DE, subjects faced the binary choice between receiving $c = 100$€ (≈$110) as payment, or saving a human life in expectation. As part of the experimental design, we predetermined this single value of $c = 100$€ as a compromise between two practical concerns: (i) c must be high enough to generate choices of both types; (ii) in contrast to *MPL*, each implemented decision has a certain cost to the experimental budget of either c or the full 350€ donation, which quickly adds up given our large subject sample.

For the MPL conditions, we used a price-list design: starting with $c = 0$€ and proceeding in 10€ increments up to $c = 200$€, subjects could indicate in each of the 21 contingent choices whether they want to save a life or take c for themselves. Each price was then equally likely to be drawn for implementation (uniform G).[7] Figures B.1 and B.2 in Appendix B provide screenshots of the decision screens.

Turning to visibility, recall that the two key forces underlying Proposition 3, namely the *discouragement* and the *cheap-act* effects, both require a non-zero level of image concerns. The former is then predicted to dominate at low values, leading to $\bar{a}^{DE}(c,\mu) > \bar{a}^{MPL}(c,\mu)$, and the latter at high values, reversing the inequality. To ensure a minimal level of image concern in both treatments, we notified subjects at the start that: (i) they were anonymously paired with another participant in the same session; and (ii) they would see, at the end of

[7] To avoid inconsistent answers, we enforced a single-switching rule.

the experiment, their own and their partner's choices displayed alongside on their screens, as would their partner. Apart from observing the partner's choices, subjects received no information about them, so that no other aspect of the dyad would influence decisions.

To keep image concerns minimal in the *Low Image* treatment ($\mu = \mu_L$), we made the payment procedure double-blind, so that not even the experimenter could link subjects' decisions to their identity. Following Barmettler et al. (2012), at the start of each session one subject was randomly designated to carry out all payments: they did not participate in the regular experiment, and thus had no knowledge about the choices. At the end, payments were stuffed into envelopes and given to the selected subject, who then handed them to those who actively participated, in an adjacent room.

The *High Image* treatment ($\mu = \mu_H$), in contrast, was designed to induce strong image concerns. Subjects were informed that upon receiving payment their choice would be observed and compared to those of their matched partner by a committee of three persons, sitting in the room where payments would take place. Both choices were projected onto a wall, and subjects had to state their choices and those of their partners aloud, using two predetermined sentences.

4.3 Procedure

697 subjects (405 female, mean age = 24.01, SD = 6.21) participated in 36 sessions at the BonnEconLab of the University of Bonn: 178 in the *MPL-Low Image* treatment, 178 in *MPL-High Image*, 165 in *DE-Low Image*, and 176 in *DE-High Image*. Subjects were recruited using the software hroot (Bock et al., 2014), and the experiment was conducted using oTree (Chen et al., 2016). Sessions lasted about 60 minutes, with a show-up fee of 12€. For each session, one matched pair of subjects was randomly drawn, and their choices implemented. Thus, in the DE treatments, each of the two either received 100€, or triggered a life-saving 350€ donation. In the *MPL* treatments, one price from the list was randomly drawn (uniformly), and the pre-stated choices of both partners for this price were implemented. Therefore, each one either triggered the donation or received up to 200€.[8]

At the beginning of each session, subjects received a verbal introduction to the experiment. In the *Low Image* treatments, the procedure ensuring anonymity was explained and demonstrated. In the *High Image* treatments, the setup was shown. Subsequently, all subjects received detailed information about tuberculosis, its effects, and treatment. The instructions also linked to a website where they were invited to confirm the validity of the information. We then introduced the charity and its working procedure, and explained our calculations

[8]This random implementation adds another layer of the cheap-talk effect, but one that affects *DE* and *MPL* in exactly the same way (formally equivalent to dividing μ by the probability of implementation), and thus leaves all comparisons between the two unaffected.

regarding the life-saving effect of the 350€ donation. Subjects then learned about their choice options and, after answering a couple of comprehension questions, made their decisions. Finally, they completed a short questionnaire and were paid in a separate room, with payment procedures depending on treatment status, as explained above.

5 Hypotheses and Results

Our outcome variable is the fraction $\bar{a}^m(c, \mu)$ of subjects who choose to save a life over receiving c, given an elicitation method $m \in \{DE, MPL\}$ and a level of visibility $\mu \in \{\mu_L, \mu_H\}$. For brevity, we will refer to $\bar{a}^m(c, \mu)$ as "total contributions".

5.1 Hypotheses

Based on Proposition 3, we state:

Hypothesis 1. *For both* DE *and* MPL*, total contributions are higher under* High Image *than under* Low Image: $\bar{a}^{DE}(c, \mu_H) > \bar{a}^{DE}(c, \mu_L)$, $\bar{a}^{MPL}(c, \mu_H) > \bar{a}^{MPL}(c, \mu_L)$.

Hypothesis 2. *Under* Low Image*, total contributions are higher under* DE *than under* MPL*:* $\bar{a}^{DE}(c, \mu_L) > \bar{a}^{MPL}(c, \mu_L)$.

Hypothesis 3. *Under* High Image*, total contributions are higher under* MPL *than under* DE*:* $\bar{a}^{DE}(c, \mu_H) < \bar{a}^{MPL}(c, \mu_H)$.

Hypothesis 1 captures the standard effect of signaling concerns. The novel ones are Hypotheses 2 and 3**,** reflecting the dominance of the *discouragement effect* at μ_L and the *cheap-act effect* at μ_H. Together, they constitute the model's distinctive crossing prediction, which we will test at $c = 100$€, as explained earlier.

5.2 Results

Hypothesis 1. Under both elicitation methods, increased visibility led to a rise in total contributions, but the magnitude was markedly different. Under *DE*, 58.8% of subjects chose to save a life in *Low Image* and 62.5% in *High Image* –a relatively small and insignificant increase ($p = 0.51$).[9] Under *MPL*, increased visibility had a much larger effect. At almost all payment levels, the fraction of subjects choosing to save a life is at least 15 pp. higher under *MPL-High Image* than under *MPL-Low Image*, resulting in significantly different distributions

[9]We follow the convention of reporting two-sided tests, even though we are testing the directed (inequality) hypotheses emanating from the model. Unless otherwise noted, we use Fisher's exact test.

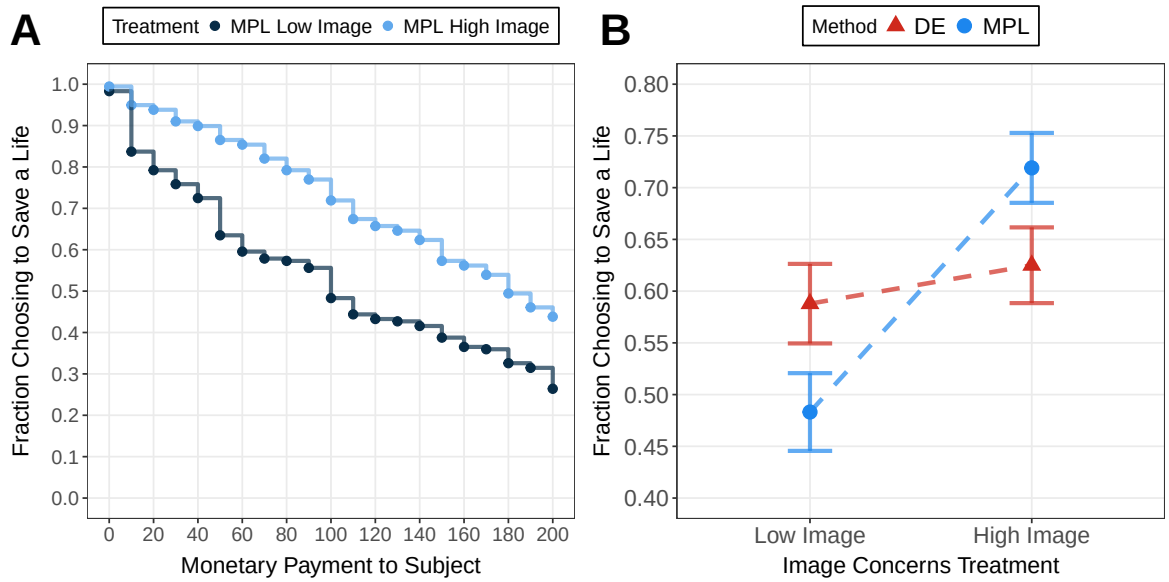

Figure 2: Main Results. Panel A displays the fraction choosing to save a life for each offered price in the MPL Low Image and MPL High Image treatments. Panel B shows the interaction effect of elicitation method and image concerns, by displaying the fractions choosing to save a life with MPL and DE, under either the Low Image or the High Image treatment. Error bars indicate the standard error of the mean.

($p < 0.001$, Kolmogorov–Smirnov test); see Panel A of Figure 2. At 100 €, contributions are 23.6 pp. and significantly higher under *High Image* than under *Low Image* ($p < 0.001$).

Hypotheses 2 and 3. Panel B of Figure 2 shows that the fractions $\bar{a}^m(100, \mu)$ choosing to save a life over 100€ clearly differ by elicitation method, with the ranking reversing between μ_L and high μ_H. Under *Low Image*, we observe $\bar{a}^{MPL}(\mu_L) < \bar{a}^{DE}(\mu_L)$, as predicted by Hypothesis 2, and consistent with the dominance of the *discouragement effect*. The difference is large, with the fraction saving a life rising from 48.3% to 58.8% between *MPL* and *DE* ($p = 0.065$). In contrast, under *High Image* we observe $\bar{a}^{MPL}(\mu_H) > \bar{a}^{DE}(\mu_H)$, in line with the *cheap-act effect* dominating, as predicted by Hypothesis 3. The difference is again about 10 percentage points, but now in the opposite direction, rising from 62.5% under DE to 71.9% under MPL ($p = 0.070$).

Table 1, Panel A regresses the probability of choosing to save a life (instead of taking 100€) on a dummy for the type of elicitation (1 for *MPL*), which yields a positive coefficient for *Low Image* in Column (1), and a negative one for *High Image* in Column (3).[10] Columns (2) and (4), show that these effects remain largely unaffected by controls for age, gender, high-school graduation grade, highest educational degree obtained so far, self-reported monthly income, and a measure of religiousness (Likert scale).

Combining Hypotheses 2 and 3 implies that the increase in contributions as image rises from μ_L to μ_H should be more pronounced for *MPL* than for *DE*. Panel B of Table 1 thus presents an OLS regression interacting *High Image* with *MPL,* using *DE-Low Image* as base-

[10]The results remain qualitatively unchanged with Probit or Logit regressions.

line; the interaction is positive and significant at the 1-percent level. In sum, our results support the key predictions emanating from Proposition 3.

Robustness Experiment. One may worry that features of the elicitations methods unrelated to image concerns might be at play in our results. Note first that these would have to generate not just different *DE* versus *MPL* contributions, but also a flipping of that gap as image rises from low to high, which seems rather implausible. Nonetheless, to rule out such confounding factors altogether, we ran the *DE* versus *MPL* treatments on another 366 subjects, with the donation replaced by a non-moral good (university shop voucher). For this "placebo", $\mu = 0$, and indeed we find no significant differences between *MPL* and *DE*: see Panel C of Table 1, and Appendix C for implementation details.

6 Conclusion

Our model and experiment show that image concerns affect the measurement of moral preferences in ways that *interact with the elicitation method*. Regardless of whether one is interested in image-inclusive preferences (for positive predictions) or in purely intrinsic ones (for normative judgements), behavior will differ between direct and price-list mechanisms. These results argue for caution in interpreting standard estimates of moral preferences from experiments and contingent-valuation surveys,[11] but also provide potential guidance for maximizing public-goods contributions and image manipulations.[12]

In particular, even purely utilitarian individuals may act, when facing *MPL*-like situations, as if deontologically motivated: refusing all proposed prices in exchange for what is perceived as having a dignity. With necessarily finite budgets, a definitive test of how many "real Kantians" there are is ultimately impossible, but our experiment provides both an upper bound and some grounds for skepticism about public positions on the subject. The former is given by the 26.4% of subjects who choose to save a life over the maximum offer of $200€$ in the Low Image *MPL* condition. The latter stems from the fact that this proportion nearly doubles to 43.82% with a mild visibility manipulation. These results can also help to account for the common resistance to estimating and using a "statistical value of life." Despite the fact that we implicitly engage in trading off costs and statistical lives all the time, explicit reference to putting a price tag on life typically produces conspicuously displayed righteous indignation (Sandel, 2012).

[11] A related point is made by Chen and Schonger (2016b) for other forms of preferences involving moral "duties".

[12] Individual WTP's, which include the value of social and self-image, are the right measures to predict, explain or alter behavior. To inform policy, however, they can substantially overstate the true social value of the public good. Thus, in our model, reputation is a positional good, the image gains and losses of contributors and non-contributors exactly offsetting each other. In general, the image game can have negative, zero, or positive sum, depending on the curvature of the reputation functional; Butera et al. (2022) find evidence for negative sum, which reinforces the previous point.

Table 1: Regression analyses of the effect of the elicitation method on prosocial behavior

Panel A:

Dependent variable:	Choice to Save a Life (vs. 100€)			
	Low Image Concerns		*High Image Concerns*	
	(1)	(2)	(3)	(4)
MPL	−0.105	−0.103	0.094	0.091
	(0.054)	(0.053)	(0.050)	(0.050)
Constant (*DE*)	0.588	0.626	0.625	0.622
	(0.038)	(0.049)	(0.037)	(0.046)
Controls		X		X
Observations	343	343	354	354

Panel B:

Dependent variable:	Choice to Save a Life (vs. 100€)	
	(1)	(2)
MPL	-0.105	-0.097
	(0.054)	(0.053)
High Image	0.037	0.052
	(0.053)	(0.052)
MPL X High Image	0.199	0.190
	(0.073)	(0.072)
Constant (*DE Low Image*)	0.588	0.595
	(0.038)	(0.044)
Controls		X
Observations	697	697

Panel C:

Dependent variable:	Choice of Voucher (vs. 10€)	
	(1)	(2)
MPL No-Image	0.045	0.051
	(0.047)	(0.047)
Constant (*DE No-Image*)	0.253	0.227
	(0.033)	(0.047)
Controls		X
Observations	366	366

The table shows OLS regression coefficients. Robust standard errors in parentheses. Controls include age, gender, income, religiousness, educational level, and high school grade.

References

Alger, Ingela, and Jörgen W. Weibull (2013). "Homo Moralis–Preference Evolution Under Incomplete Information and Assortative Matching". *Econometrica* 81 (6): 2269–302. [4]

Ambuehl, Sandro (2017). "An Offer You Can't Refuse? Incentives Change How We Inform Ourselves and What We Believe". *CESifo Working Paper No. 6296*, [4]

Andreoni, James (1989). "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence". *Journal of Political Economy* 97 (6): 1447–58. [5]

Andreoni, James (1990). "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow". *Economic Journal* 100 (401): 464–77. [5]

Ariely, Dan, Anat Bracha, and Stephan Meier (2009). "Doing Good or Doing Well?: Imge Motivation adn Monetary Incentives in Behaving Proscially". *American Economic Review* 99 (1): 544–55. [4]

Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack (2014). "No margin, no mission? A field experiment on incentives for public service delivery". *Journal of Public Economics* 120: 1–17. [4]

Baillon, Aurelien, Yoram Halevy, and Chen Li (2022). "Randomize at your own Risk: on the Observability of Ambiguity Aversion". *Econometrica*, (forthcoming): [4]

Barmettler, Franziska, Ernst Fehr, and Christian Zehnder (2012). "Big experimenter is watching you! Anonymity and prosocial behavior in the laboratory". *Games and Economic Behavior* 75 (1): 17–34. [12]

Bartling, Björn, Vanessa Valero, Roberto Weber, and Yao Lan (2020). "Public Discourse and Socially Responsible Market Behavior". *Working Paper*, [4]

Bénabou, Roland, Armin Falk, and Jean Tirole (2020). "Narratives, Imperatives, and Moral Persuasion". *Working Paper*, [4]

Bénabou, Roland, and Jean Tirole (2006). "Incentives and prosocial behavior". *American Economic Review* 96 (5): 1652–78. [4]

Bénabou, Roland, and Jean Tirole (2011a). "Identity, morals, and taboos: Beliefs as assets". *Quarterly Journal of Economics* 126 (2): 805–55. [4]

Bénabou, Roland, and Jean Tirole (2011b). "Laws and Norms". *NBER Working Paper 17579*, [4]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "Hroot: Hamburg Registration and Organization Online Tool". *European Economic Review* 71: 117–20. [12]

Bos, Olivier, and Martin Pollrich (2020). "Optimal Auctions with Signaling Bidders". *Working Paper*, [4]

Bos, Olivier, and Tom Truyts (2022). "Entry in First-Price Auctions with Signaling". *Working Paper*, [4]

Brandts, Jordi, and Gary Charness (2011). "The strategy versus the direct-response method: A first survey of experimental comparisons". *Experimental Economics* 14 (3): 375–98. [4]

Brekke, Kjell A., Snorre Kverndokk, and Karine Nyborg (2003). "An economic model of moral motivation". *Journal of Public Economics* 87 (9-10): 1967–83. [4]

Brock, J. Michelle, Andreas Lange, and Erkut Y. Ozbay (2013). "Dictating the Risk: Experimental Evidence on Giving in Risky Environments". *American Economic Review* 103 (1): 415–37. [5]

Bursztyn, Leonardo, Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao (2018). "Status Goods: Experimental evidence from platinum credit cards". *Quarterly Journal of Economics* 133 (3): 1561–95. [4]

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky (2022). "Measuring the Welfare Effects of Shame and Pride". *American Economic Review* 112 (1): 122–68. [15]

Charness, Gary, Uri Gneezy, and Brianna Halladay (2016). "Experimental methods: Pay one or pay all". *Journal of Economic Behavior and Organization* 131: 141–50. [4]

Charness, Gary, Uri Gneezy, and Alex Imas (2013). "Experimental methods: Eliciting risk preferences". *Journal of Economic Behavior and Organization* 87: 43–51. [4]

Chen, Daniel L., and Martin Schonger (2016a). "A Theory of Experiments: Invariance of Equilibrium to the Strategy Method of Elicitation and Implications for Social Preferences". *TSE Working Paper, no. 16-724*, [4]

Chen, Daniel L., and Martin Schonger (2016b). "Social preferences or sacred values? Theory and evidence of deontological motivations". *TSE Working Paper, no. 16-714*, [5, 15]

Chen, Daniel L., Martin Schonger, and Chris Wickens (2016). "oTree-An open-source platform for laboratory, online, and field experiments". *Journal of Behavioral and Experimental Finance* 9: 88–97. [12]

Cohen, Jonathan, Keith Marzilli Ericson, David Laibson, and John Myles White (2020). "Measuring time preferences". *Journal of Economic Literature* 58 (2): 299–347. [4]

Cox, James C., Vjollca Sadiraj, and Ulrich Schmidt (2015). "Paradoxes and mechanisms for choice under risk". *Experimental Economics* 18 (2): 215–50. [4]

DellaVigna, Stefano, John A. List, and Ulrike Malmendier (2012). "Testing for altruism and social pressure in charitable giving". *Quarterly Journal of Economics* 127 (1): 1–56. [4]

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao (2022). "Estimating Social Preferences and Gift Exchange with a Piece-Rate Design". *American Economic Review* 112 (3): 1038–74. [5]

Elias, Julio, Mario Macis, and Nicola Lacetera (2016). "Efficiency-Morality Trade-Offs in Repugnant Transactions: A Choice Experiment". *IZA Discussion Paper No. 10187*, [4]

Ellingsen, Tore, and Magnus Johannesson (2008). "Pride and Predjudice: The Human Side of Incentive Theory". *American Economic Review* 98 (3): 1–40. [4]

Exley, Christine L. (2016). "Excusing selfishness in charitable giving: The role of risk". *Review of Economic Studies* 83 (2): 587–628. [4]

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2018). "Global Evidence on Economic Preferences". *Quarterly Journal of Economics* 133 (4): 1645–92. [4]

Falk, Armin, and Thomas Graeber (2020). "Delayed Negative Effects of Prosocial Spending on Happiness". *Proceedings of the National Academy of Sciences* 117 (12): 6463–68. [10]

Falk, Armin, Thomas Neuber, and Nora Szech (2020). "Diffusion of being pivotal and immoral outcomes". *Review of Economic Studies* 87 (5): 2205–29. [4, 5]

Falk, Armin, and Nora Szech (2013). "Morals and markets". *Science* 340 (6133): 707–11. [4]

Feddersen, Timothy, Sean Gailmard, and Alvaro Sandroni (2009). "Moral bias in large elections: Theory and experimental evidence". *American Political Science Review* 103 (2): 175–92. [5]

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. [4]

Giovannoni, Francesco, and Miltiadis Makris (2014). "Reputational bidding". *International Economic Review* 55 (3): 693–710. [4]

Gneezy, Uri, Elizabeth A. Keenan, and Ayelet Gneezy (2014). "Avoiding overhead aversion in charity". *Science* 346 (6209): 632–35. [5]

Goeree, Jacob K. (2003). "Bidding for the future: Signaling in auctions with an aftermarket". *Journal of Economic Theory* 108 (2): 345–64. [4]

Goeree, Jacob K., Charles A. Holt, and Susan K. Laury (2002). "Private costs and public benefits: Unraveling the effects of altruism and noisy behavior". *Journal of Public Economics* 83 (2): 255–76. [4]

Grossman, Zachary (2015). "Self-signaling and social-signaling in giving". *Journal of Economic Behavior and Organization* 117: 26–39. [3]

Grossman, Zachary, and Joël J. van der Weele (2017). "Self-image and willful ignorance in social decisions". *Journal of the European Economic Association* 15 (1): 173–217. [4]

Kagel, John H., and Alvin E. Roth (1995). *The Handbook of Experimental Economics*. Princeton: Princeton University Press. [4]

Kant, Immanuel (1785). *Grundlegung zur Metaphysik der Sitten*. J. F. Hartknoch. [1]

Kolappan, C., R. Subramani, V. Kumaraswami, T. Santha, and P. R. Narayanan (2008). "Excess mortality and risk factors for mortality among a cohort of TB patients from rural south India". *International Journal of Tuberculosis and Lung Disease* 12 (1): 81–86. [11]

Lacetera, Nicola, Mario Macis, and Robert Slonim (2012). "Will there be blood? Incentives and displacement effects in pro-social behavior". *American Economic Journal: Economic Policy* 4 (1): 186–223. [4]

Roemer, John E. (2010). "Kantian equilibrium". *Scandinavian Journal of Economics* 112 (1): 1–24. [4]

Sandel, Michael J. (2012). *What Money Can't Buy: The Moral Limits of Markets*. Farrar, Straus, and Giroux. [15]

Selten, Reinhard (1967). "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments". In *Beiträge zur experimentellen Wirtschaftsforschung*. H. Sauermann, ed. Tübingen: Mohr, 136–68. [4]

Straetemans, Masja, Philippe Glaziou, Ana L. Bierrenbach, Charalambos Sismanidis, and Marieke J. van der Werf (2011). "Assessing tuberculosis case fatality ratio: A meta-analysis". *PLoS ONE* 6 (6): [11]

Tiemersma, Edine W., Marieke J. van der Werf, Martien W. Borgdorff, Brian G. Williams, and Nico J.D. Nagelkerke (2011). "Natural history of tuberculosis: Duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: A systematic review". *PLoS ONE* 6 (4): [11]

ONLINE APPENDIX

A Proofs

Proof of Proposition 1. From conditions (2)-(4), it is straightforward to characterize the regions in which each possible equilibrium exists:

(P_0) Pooling at $a_H = a_L = 0$, sustained by out-of equilibrium beliefs v_H following $a = 1$ (by the D1 criterion), is an equilibrium if and only if $c \geq c_H^{DE}$. In the intermediate range where

$$\bar{c}_L^{DE} = v_L e + \mu(v_H - v_L) \leq c \leq v_H e + \mu(v_H - v_L) \equiv \bar{c}_H^{DE},$$

it coexists with a separating equilibrium in which the high type contributes, and a mixed-strategy one in-between. Intuitively, if the high type is expected to abstain there is less stigma from doing so, which in turn reduces his incentive to contribute. The no-contribution pooling equilibrium is better for both types (Pareto dominant), however: the separating one yields only $\mu \cdot v_L < \mu\bar{v}$ for the low type and $v_H e - c/\beta + \mu v_H \leq \mu\bar{v}$ for the high type. Whenever such multiplicity arises, we shall therefore select the $a_H = a_L = 0$ equilbrium.

(P_1) Pooling at $a_H = a_L = 1$, sustained by out-of equilibrium beliefs v_L following $a = 0$ (by the D1 criterion), is an equilibrium if and only if $c \leq \underline{c}_L^{DE}$.

(S) Separation, namely $a_H = 1 - a_L = 1$, is is an equilibrium if and only if $\bar{c}_L^{DE} \leq c \leq \bar{c}_H^{DE}$, where $\bar{c}_H^{DE} > c_H^{DE}$ is defined by $v_H e - c_H^{DE} + \mu\left(v_H - \bar{v}\right) \equiv 0$.

(SS_1) Semi-separation with $0 < a_L < 1 = a_H$, and beliefs $\hat{v} \in (v_L, \bar{v})$ following $a = 1$, is an equilibrium if and only if $\underline{c}_L^{DE} < c < \bar{c}_L^{DE}$. The low type's mixed strategy $a_L(c) \in (0, 1)$ is then given by combining the indifference condition and $v_L e - c + \mu(\hat{v}(a_L) - v_L) = 0$ and the Bayesian posterior $\hat{v}(c) = \left[\rho v_H + (1-\rho)a_L v_L\right] / \left[\rho v + (1-\rho)a_L\right]$, which leads to:

$$v_L e - c + \frac{\mu\rho(v_H - v_L)}{\rho + (1-\rho)a_L(c)} \equiv 0, \tag{7}$$

so that $a_L(c)$ is decreasing in c, and the reputation $\hat{v}(c)$ following $a = 1$ conversely increasing.

(SS_0) Semi-separation with $0 = a_L < a_H < 1$, and beliefs $\hat{v} \in (\bar{v}, v_H)$ following $a = 0$, is an equilibrium if and only if $c_H^{DE} < c < \bar{c}_H^{DE}$. It thus always coexists with P_0, and is always dominated by it.

These results jointly imply that:

(a) If $\underline{c}_L^{DE} < \bar{c}_L^{DE} < c_H^{DE}$, the unique equilibrium is P_1 below the first cutoff, SS_1 between the first and second, and S between the second and third. Above the third, the dominant

equilibrium is P_0.

(b) If $\underline{c}_L^{DE} < c_H^{DE} < \bar{c}_L^{DE}$ (where the second inequality means that $\mu\rho > e$), the unique equilibrium is P_1 below the first cutoff, and SS_1 between the first and second; above that, the dominant equilibrium is P_0.

(b) If $c_H^{DE} < \underline{c}_L^{DE} < \bar{c}_L^{DE}$ (where the first inequality means that $\mu(2\rho - 1) > e$), the unique equilibrium is P_1 below the first cutoff, and above it the dominant equilibrium is P_0. ■

Proof of Proposition 2. The proof of existence is standard. For example, for a separating equilibrium to obtain, it must be: that (i) the low type obtains his symmetric-information allocation (otherwise, he would be better off selecting $c_L^{MPL} = v_L e$, , and obtain reputation equal to at least v_L anyway), and (ii) the low type does not want to mimic the high type: $\mu(v_H - v_L) \leq L(c_H^{MPL})$ and $c_H^{MP} < c_{max}$. It is easily verified that the proposed strategies satisfy these conditions. The conditions for semi-separating and pooling equilibria are also standard, and readily verified.

The proposed equilibrium is not unique in the absence of refinement, however. For example, there is a pooling equilibrium at $c^{MPL} = v_H e < c_{max}$, provided that $\mu(\bar{v} - v_L) \geq L(v_H e)$, sustained by out-of-equilibrium beliefs equal to v_L whenever the declared price differs from $v_L e$. To eliminate such multiplicity, note that equilibrium sorting implies monotonicity, so there is at most one pooling price that can be chosen with positive probability by both types. Let c^* denote this pooling price (any other price claimed by the high type exceeds c^*, and any other price claimed by the low type lies below c^*). Let $\hat{v}(c)$ denote the mean belief following price c. Applying D1, consider a deviation to price $c + \varepsilon$, for $\varepsilon > 0$ arbitrarily small, together with the set of belief responses that raise the two types' utility relative to their equilibrium utility:

$$\hat{V}_L \equiv \{\hat{v}(c^* + \varepsilon) \mid \mu\left[\hat{v}(c^* + \varepsilon) - \hat{v}(c^*)\right] > L_L(c^* + \varepsilon) - L_L(c^* + \varepsilon)\},$$
$$\hat{V}_H \equiv \{\hat{v}(c^* + \varepsilon) \mid \mu\left[\hat{v}(c^* + \varepsilon) - \hat{v}(c^*)\right] > L_H(c^* + \varepsilon) - L_H(c^* + \varepsilon)\}.$$

Clearly $V_L \subset V_H$, so a deviation to $c^* + \varepsilon$ must, by D1, induce a probability-one belief on v_H and therefore the only equilibrium pooling possibility is at price $c = c_{max}$. This, in turn, implies that the equilibrium must take one of the three forms described in the proposition, and because it is obtained on disjoint sets of parameters, the equilibrium is unique under D1. ■

B Decision Screens

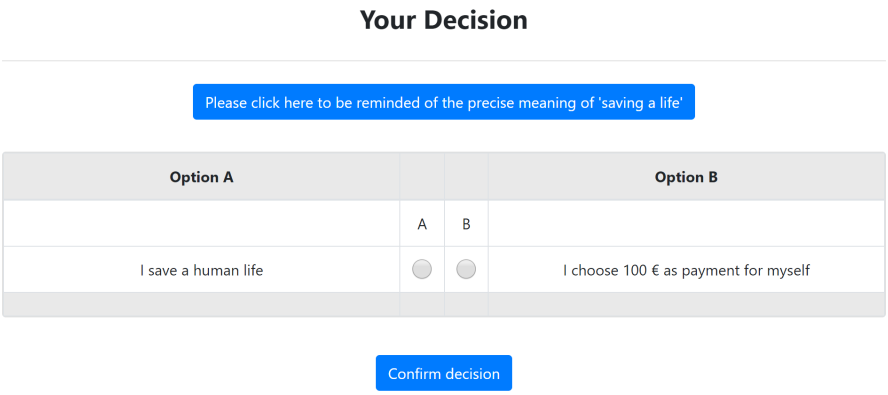

Figure B.1: Decision Screen DE

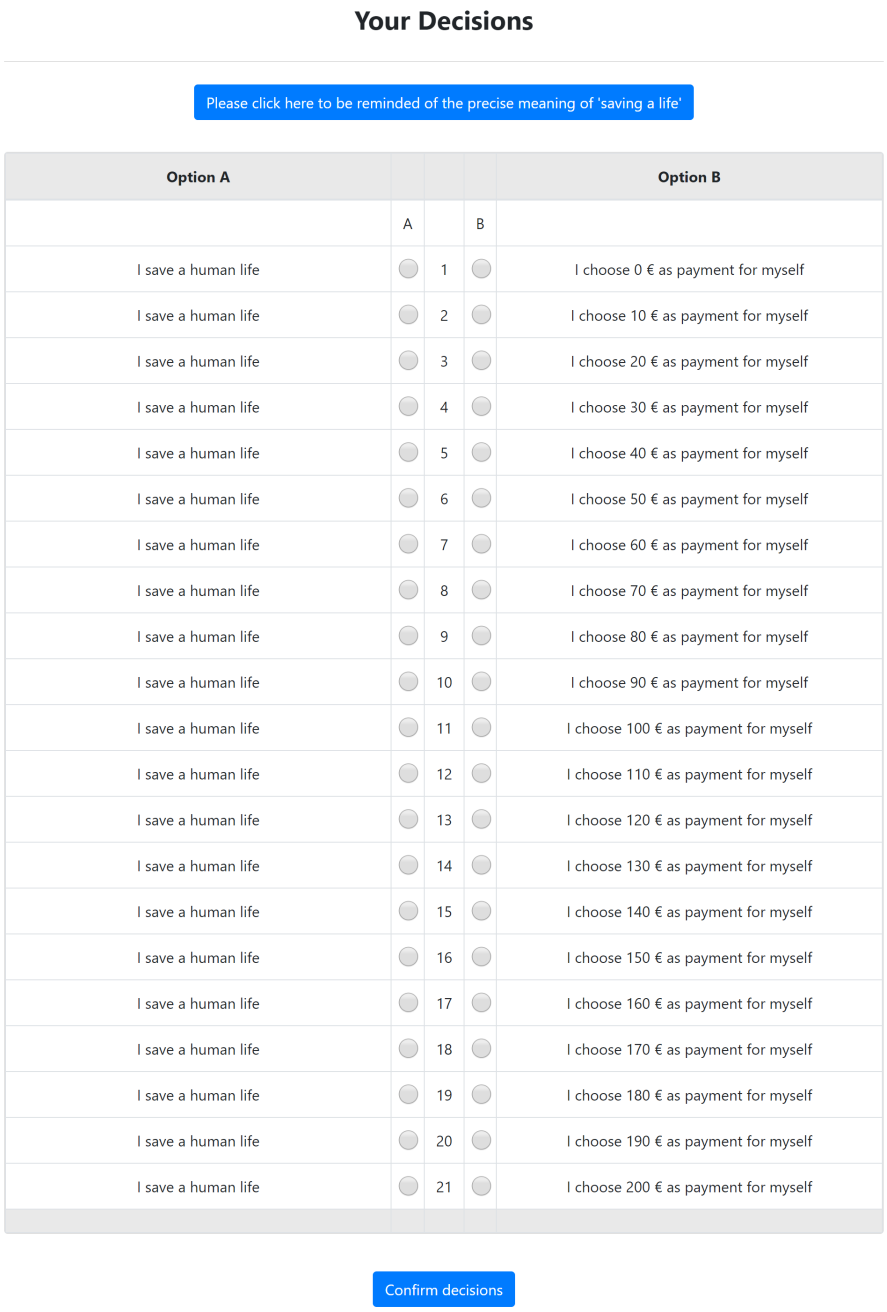

Figure B.2: Decision Screen MPL

C Robustness Experiment

In the main experiment, we showed how image concerns lead to differences in moral behavior between elicitation methods. One concern is that there are factors present in our experiment that lead to differences between DE and MPL independent of image concerns.

In particular, the previous literature has identified two main factors that could potentially confound the comparison between the two elicitation methods in our case.

First, in our experiment, only a subset of subjects had their decision implemented for real. In the MPL treatments, another randomization takes place, which is absent in DE: if selected for payout, one decision of the price list is randomly selected. If subjects violate the independence axiom and view these two randomization processes not separately but rather as a meta-lottery, this could potentially affect the comparison. This issue is also present in the many experiments that study decisions over lotteries and pay only one lottery out for real. In this context, it is usually assumed that subjects evaluate the different random processes in isolation, an assumption that has been repeatedly validated empirically[13]. It is natural to assume that subjects also perceive the two processes in isolation in our experiment since they were introduced and explained at two different points in the instructions.

The second factor is the so-called compromise effect (Andersen et al., 2006; Birnbaum, 1992; Simonson, 1989). When presenting a price list, the focus lies perceptually on the center. This in turn could change the attractiveness of the options appearing in the middle of the price list, biasing answers away from the subject's true valuations. To control for this effect, we carefully selected the DE value to correspond to the value precisely in the middle of the price list in the MPL treatments. As such, it seems unlikely that differences in perceptions could explain discrepancies between the elicitation methods.

Therefore, we would not expect differences between DE and MPL in our experiment once image concerns are absent. Nevertheless, in order to document this empirically, we conducted a robustness experiment, which is explained next.

C.1 Setup and Treatments

For the robustness experiment, we used a good that is unrelated to prosocial and moral considerations, so that image concerns are plausibly absent. For this non-moral good, we chose a 35 € voucher for the University of Bonn's online shop. With the voucher, subjects can buy sweatshirts, T-shirts, and accessories related to the university. The voucher cannot be returned and is only valid for purchases in the shop. There were two between-subject treatments: *DE No-Image* and *MPL No-Image*. In the former, subjects could choose between 10 € and the voucher, while in the latter they faced a price list from 0€ to 20€ in 1€ increments. Note that this closely mimics the decisions in the main experiment. The only difference is that all values are divided by 10. As in the main experiment, subjects were paired with another subject, and only a subset of subjects had their choices implemented for real.

[13] See e.g., Starmer and Sugden (1991), Cubitt, Starmer and Sugden (1998) and Hey and Lee (2005).

Accordingly, instructions for the decisions were identical, with the sole difference being that descriptions related to the saving a life paradigm were replaced with descriptions of the voucher. Consequently, any factors influencing the comparison between *DE* and *MPL* in the main experiment should also manifest in the robustness experiment.

C.2 Procedure

Subjects were recruited from the same subject pool as the main experiment, with the restriction that they had not previously participated in the main experiment. The experiment was conducted as a virtual lab experiment since in-person lab sessions were not possible due to the ongoing Covid-19 pandemic. That is, the experiment started and ended at a prespecified date and time, and the experimenter was available during the experiment in case of problems.

In total, 366 subjects (227 female, mean age 26.88, SD 7.87) took part, 188 in the *MPL No-Image*, and 178 in the *DE No-Image* treatment, respectively. The experiment lasted on average 13 minutes, for which the subjects received a show-up fee of 3€. Subjects were grouped in virtual sessions consisting of roughly 24 subjects, and one pair was randomly selected for payout out of each virtual session. Exactly as in the main experiment, for these two subjects, either their DE decision was implemented or a randomly chosen decision from the MPL list.

C.3 Results

Assessing subjects' general valuation of the voucher, we observe considerable variation in switching behavior in the *MPL No-Image* treatment. In total, 76% had an interior switching value, meaning they preferred the voucher in the initial decision but switched to preferring the monetary value at some point. The variation compares quite favorably to the *MPL-Low Image* treatment, where this was the case for 72% of subjects. Comparing the choice at 10 € in *MPL No-Image* with *DE No-Image*, we find that 29.8% choose the voucher in *MPL* and 25.3% in *DE*. This difference is small in magnitude and not statistically significant ($p = 0.35$; two-sided Fisher's exact test). It is also in the opposite direction of what we find in the main experiment for the *Low Image* case, which is the natural comparison. Table C.1 replicates this null result in an OLS-regression, with column (2) using the same variables as control variables as in the main experiment, compare Table 1, columns (2) and (4). Thus, we do not observe any meaningful differences between the two elicitation methods in our setting once image concerns are removed.

Table C.1: Regression analyses of the effect of the elicitation method on voucher choice

Dependent variable:	Choice of Voucher (vs. 10€)	
	(1)	(2)
MPL No-Image	0.045	0.051
	(0.047)	(0.047)
Constant (*DE No-Image*)	0.253	0.227
	(0.033)	(0.047)
Controls		X
Observations	366	366
R^2	0.003	0.039

The table shows OLS regression coefficients. Robust standard errors in parentheses. Controls include age, gender, income, religiousness, educational level, and high school grade.

References

Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabet Rutström (2006). "Elicitation using multiple price list formats ". Experimental Economics, 9(4), 383-405

Birnbaum, Michael H. (1992). "Violations of Monotonicity and Contextual Effects in Choice-Based Certainty Equivalents ". Psychological Science, 3(5), 310-315.

Simonson, Itamar (1989). "Choice Based on Reasons: The Case of Attraction and Compromise Effects ". Journal of Consumer Research, 16(2), 158.

Starmer, Chris, and Robert Sugden (1991), "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation "American Economic Review 81, 971-978.

Cubitt, Robin P., Chris Starmer, and Robert Sugden (1998). "On the Validity of the Random Lottery Incentive Mechanism "Experimental Economics 1, 115-132

Hey, John D., and Jinkwon Lee (2005), "Do Subjects Separate (or Are They Sophisticated)? "Experimental Economics 8, 233-265.

D Instructions

B.1 Announcement by the Experimenter

The following text was read aloud by the experimenter after all subjects were placed in their cubicles, establishing common knowledge among all subjects of a session. The content depended on the image treatment.

B.1.1 Treatment *Low Image*

Welcome to today's study. In today's study, you will make decisions on a computer. These decisions will take place under complete anonymity. To ensure this, we will now apply the following procedure: You should all have two notes with your cubicle number in front of you. We will soon collect one of the two notes and randomly draw one out of all collected. The person in the drawn cubicle is responsible for the payment in today's study. At the end of the study, we prepare sealed envelopes with your payments. Those envelopes are then passed to the soon to be randomly drawn person, who will hand them out to each of you sequentially in the adjacent room. The envelopes are designed so that you cannot see the contents from the outside, i.e., not on weight or similar clues. Hence at no time can there be a connection drawn between your payment and your decisions. Please hold now one of the notes with your cubicle number onto out of your cubicle. (Responsible person is drawn and placed in the adjacent room) The study will begin shortly. If you have at any time have questions, just hold your hand out of the cubicle.

B.1.2 Treatment *High Image*

Welcome to today's study. In today's study, you will make decisions on your computer. Your decisions will subsequently be evaluated by a committee consisting of three students from the University of Bonn. For this, after you have made your decisions, you will go to the adjacent room, where your decisions will be projected on a wall with a projector. You will then briefly communicate your decisions to the committee, and the committee will evaluate them. Afterward, you will receive the result of the evaluation. Detailed information about your decisions, the committee, and the evaluation will be given to you at the appropriate time on your computer. The study will begin shortly. If you have at any time have questions, just hold your hand out of the cubicle.

B.1.3 Further Procedure

After the text was read aloud, in the *Low Image* conditions the experimenter then collected one note from each subject indicating their respective cabin number. All notes were thrown into a bag, and one was drawn in front of all participants to make clear that the person responsible for the payment procedure was a randomly determined participant. In the *High Image* conditions, subjects were shown the adjacent room and the setup with the committee, which consisted of student research assistants. The members of the committee did not interact with the subjects in any way.

B.2 Introduction

All further instructions were displayed on the subjects screens. The following introduction was the same for all treatments.

B.2.1 Welcome to the study

Welcome, and thank you for your interest in today's study!

For your participation, you will receive a fixed payment of 12€ given to you at the end. In this study, you will make decisions on the computer. Depending on how you choose, you can earn additional money.

During the entire study, communication between participants is prohibited. Please turn off your phone so that other participants are not disturbed. Please only use the designated functions on the computer and make the entries with the mouse and keyboard. If you, at some point, have questions, please make a hand signal. Your question will be answered at your seat.

On the next screens, you will receive specific information about participation in this study. To proceed, click "Next".

B.2.2 Your Partner

As part of this experiment, a partner has been assigned to you. This partner is a participant in today's experiment, just like you. He or she was randomly assigned to you and will receive the same instructions as you.

In today's experiment, you and your partner will both receive the exact same information and subsequently face the exact same decisions. These decisions have certain consequences, which will be described in detail later.

At the end of today's experiment, one pair is randomly drawn from all participants in today's experiment. Only the decisions of this pair will be implemented, as described in the instructions. Please note: The random draw of a pair is completely independent of the participants' decisions. Each pair has the same probability of being drawn. Since your decision can be actually implemented for real, you should think carefully about how you will decide in the experiment.

B.2.3 Information about Tuberculosis

What follows is important information that is relevant to the decisions you will later be asked to make. It concerns the illness tuberculosis and its possible treatment. Please read through all the information carefully.

What is Tuberculosis?

Tuberculosis – also called Phthisis or White Death – is an infectious disease, which is caused by bacteria. Roughly one-third of all humans are infected with the pathogen of Tuberculosis. Active Tuberculosis breaks out among 5 to 10% of all those infected. Tuberculosis is primarily airborne. This is also why quick treatment is necessary.

Tuberculosis patients often suffer from very unspecific symptoms like fatigue, the feeling of weakness, lack of appetite, and weight loss. At an advanced stage of lung tuberculosis, the patient coughs up blood, leading to the so-called rush of blood. Without treatment, a person with Tuberculosis dies with a probability of 43%.

How prevalent is Tuberculosis?

In the year 2014, 6 million people have been recorded as falling ill with active Tuberculosis. Almost 1.5 million people die of Tuberculosis each year. This means more deaths due to Tuberculosis than due to HIV, malaria, or any other infectious disease.

Is tuberculosis curable?

According to the World Health Organization (WHO), the United Nations agency for international public health, "tuberculosis is preventable and curable". Treatment takes place by taking antibiotics several times a week over a period of 6 months. It is important to take the medication consistently. Since 2000, an estimated 53 million lives have been saved through effective diagnosis and treatment of tuberculosis.

The success rate of treatment for a new infection is usually over 85%.

The preceding figures and information have been provided by the WHO and are freely available. Click here for more details.

Operation ASHA

Figure B.1: Typical appearance of a tuberculosis patient.

Operation ASHA is a charity organization specialized since 2005 on treating Tuberculosis in disadvantaged communities. The work of *Operation ASHA* is based on the insight that the biggest obstacle for the treatment of Tuberculosis is the interruption of the necessary 6-month-long regular intake of medication.

For a successful treatment, the patient has to come to a medical facility twice a week – more than 60 times in total – to take the medication. Interruption or termination of the treatment is fatal because this strongly enhances the development of a drug-resistant form of Tuberculosis. This form of Tuberculosis is much more difficult to treat and almost always leads to death.

The Concept of *Operation ASHA*

To overcome this problem, *Operation ASHA* developed a concept that guarantees regular treatment through immediate spatial proximity to the patient. A possible non-adherence is additionally prevented by visiting the patient at home.

By now, *Operation ASHA* runs more than 360 treatment centers, almost all of which are located in the poorer regions of India. More than 60,000 sick persons have been identified and treated that way.

Operation ASHA is an internationally recognized organization, and its success has been covered by the New York Times, BBC, and Deutsche Welle, for example. The MIT and the University College London have already conducted research projects about the fight against Tuberculosis in cooperation with *Operation ASHA*. The treatment method employed by *Operation ASHA* is described by the World Health Organization (WHO) as “highly efficient and cost-effective”.

The Impact of a Donation to *Operation ASHA*

It is now possible to save people from death by Tuberculosis by donating to *Operation ASHA*.

Figure B.2: A worker from Operation ASHA delivers medication to a tuberculosis patient.

To save a person's life means here to successfully cure a person with Tuberculosis, who otherwise would die because of the Tuberculosis. A donation of 350€ ensures that at least one human life can be expected to be saved. The information used to calculate the donation amount is obtained from public statements from the World Health Organization (WHO), peer-reviewed research studies, Indian Government statistics, and published figures from *Operation ASHA*.

In the calculation, information was conservatively interpreted, or a pessimistic number was used so that the donation amount of 350€ is in the case of doubt higher than the actual costs to save a human life. In addition, in the calculation of the treatment success rate of *Operation ASHA*, the mortality rate for alternative treatment by the state tuberculosis program in India and the different detection rates for new cases of Tuberculosis are included.

In the context of this study, an agreement made with *Operation ASHA* will ensure that 100% of the donation will be used exclusively for the diagnosis and treatment of tuberculosis patients. This means that every Euro of the donation amount goes directly to saving human lives, and no other costs will be covered. Based on a very high number of cases, the contribution of a donation of 350€ can be simplified visualized as follows:

With a donation of 350€ 5 additional patients infected with Tuberculosis can be treated through *Operation ASHA*.

If these 5 persons are not treated through *Operation ASHA*, it is expected that one patient will die.

If, through the donation of 350€ all 5 patients are treated, it is expected that no patient will die.

Based on this experience, this means that through a donation of 350€ the life of a human will be saved. The relationship between a donation of 350€ and the saving of a human is

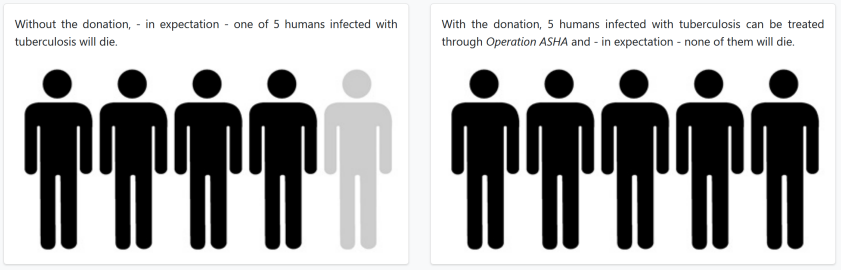

Figure B.3: Relationship between the donation and the saving of a life

illustrated in the following graphic: [Figure B.3 here]

Summary

Tuberculosis is a worldwide common bacterial infectious disease. The success rate of medical treatment of a new disease is very high. Nevertheless, close to 1.5 million people die every year from Tuberculosis. The biggest obstacle to the curing of Tuberculosis is the potential stopping of continuous treatment with antibiotics. The concept of *Operation ASHA* is therefore based on the immediate proximity to the patient as well as the control and recording of the regular intake of medication. Through a donation of 350€ to *Operation ASHA*, a life will be saved.

How is the donation connected to the saving of a life?

The donation of 350€ already accounts for the fact that someone inflicted with the illness could have survived without treatment by *Operation ASHA*; i.e., instead of through *Operation ASHA*, they could have received treatment through other actors (such as the public health system). The amount is, therefore, sufficient for the diagnosis and complete treatment of multiple sufferers.

What does it mean to “save a life”?

To save a life means here the successful curing of a person suffering from Tuberculosis, who otherwise would die because of the Tuberculosis. In particular, this means that the amount of the donation is sufficient to identify and cure so many tuberculosis patients that there is at least one person among them who otherwise could be anticipated to have died of Tuberculosis.

Note

Click on “Next”once you have finished carefully reading through the information.

You can only click on the button “Next”once you have spent at least 5 minutes on the tabs of this page.

B.3 Treatment *DE Low Image*

B.3.1 Your Decision

You will soon have the possibility to choose between two options: option A and option B. Both options are as follows:

Option A

Option A: I save a human life. By choosing option A, you save a human life. Specifically, by choosing option A, you instigate a donation of 350,00€ that will ensure that at least one person is saved from death by Tuberculosis, just as described before. If you choose option A, you will not receive an additional payment.

Option B

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment. In addition, the absence of your donation will cause the death of a human life.

Additional Payment

Before today's experiment, various amounts between 0€ and 200€ were taken into account for the amount of money you will receive when choosing option B, from which 100€ was selected. Your partner sees exactly the same options as you and makes a decision just like you. So your partner also decides between option A (saving a human life) and option B (keeping 100€ to himself).

Summary

You will decide on the next page of the screen by choosing between option A and option B. By choosing option A, you save a human life. By choosing option B, you receive an additional payment of 100€. On the next page, you will receive details about the payment procedure.

B.3.2 Further Procedure

After you confirmed your decision on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. Thereafter, some additional questions will follow. After you have answered these questions, you will get the screenshot with the decision of your partner displayed, and your partner will get the screenshot with your decision. You will not receive any further information about your partner, and your partner will not receive any further information about you.

After you received the screenshot, please remain seated until you are called with your cabin number. Then you can go into the adjacent room to pick up your compensation for today's experiment. You will be called one by one so that there is no contact with other participants of the experiment.

Who will be in the adjacent room?

In the adjacent room, you will find the participant who was randomly selected from all participants at the start of the study.

How do you receive your payment?

This participant will give you a sealed envelope with your payment. The selected participant has already received the envelope sealed. Since this participant is only responsible for the payment, this participant has not completed the study and therefore has no knowledge of the decisions to be made. Therefore, this participant does not know what you chose, how you decided, or how much money you received, exactly as explained at the beginning of the study. By handing in your note with your cabin number, you will receive the envelope intended for you.

Data protection

The subsequent analysis of all data is carried out anonymously so that your decision can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decision is only used for anonymized data analysis.

Please note:

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to experimente@briq-institute.org after the study.

B.4 Treatment *DE High Image*

B.4.1 Your Decision

You will soon have the possibility to choose between two options: option A and option B. Both options are as follows:

Option A

Option A: I save a human life. By choosing option A, you save a human life. Specifically, by choosing option A, you instigate a donation of 350,00€ that will ensure that at least one person is saved from death by Tuberculosis, just as described before. If you choose option A, you will not receive an additional payment.

Option B

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment. In addition, the absence of your donation will cause the death of a human life.

Additional Payment

Before today's experiment, various amounts between 0€ and 200€ were taken into account for the amount of money you will receive when choosing option B, from which 100€ was selected. Your partner sees exactly the same options as you and makes a decision just like you. So your partner also decides between option A (saving a human life) and option B (keeping 100€ to himself).

Summary

You will decide on the next page of the screen by choosing between option A and option B. By choosing option A, you save a human life. By choosing option B, you receive an additional payment of 100€. On the next page, you will receive details about the payment procedure.

B.4.2 Further Procedure

After you confirmed your decision on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. Thereafter, some additional questions will follow. After you have answered these questions, you will get the screenshot with the decision of your partner displayed, and your partner will get the screenshot with your decision. You will not receive any further information about your partner, and your partner will not receive any further information about you.

After you received the screenshot, please remain seated until you are called with your cabin number. Then you can go into the adjacent room to pick up your compensation for today's experiment. You will be called one by one so that there is no contact with other participants of the experiment.

Who will be in the adjacent room?

In the adjacent room, you will find a person who will make the payment. As mentioned before, there will also be a committee consisting of three other persons in the adjacent room. These three persons are students of the University of Bonn and were specially selected for this task.

What is the task of the committee?

The task of these three persons is to assess the decision you and your partner have taken. Specifically, the assessment is about how moral your behavior and the behavior of your partner is. Apart from the assessment, the three persons will not interact with you (or with your partner) in any way, and the rating will not influence at all the consequences of your decisions or your payment.

What information does the committee receive?

In order for the committee to make the assessment, the two screenshots of the decision of you and your partner are projected side by side onto the room's wall using a projector, visibly for all persons in the adjacent room. You are identified by (and only by) your cabin number. For better identification, based on your decision and that of your partner, you must also say the following two sentences aloud. The first sentence refers to your decision, the second sentence to your partner's decision.

Sentence 1 in case you chose option A: "I have decided not to take 100€ as payment for myself and instead decided to save a human life."Sentence 1 in case you chose option B: "I have decided to take 100€ as payment for myself instead of saving a human life."

Sentence 2 in case your partner has chosen option A: "My partner has decided not to take 100€ as payment for himself and instead decided to save a human life."Sentence 2 in case your partner has chosen option A: "My partner has decided to take 100€ as payment for himself instead of saving a human life."

In summary, you have to say two sentences, and the following information will be visible to everyone in the room:

- The decision you and your partner faced.
- Which option you and your partner have chosen. This means it is displayed whether you chose to save a human life or the additional payment of 100€ and whether your partner chose to save a human life or the additional payment of 100€.

How does the assessment work?

The committee will assess your decision using a scale. For this, each one of the three persons of the committee selects one of the following five values:

1 - very immoral 2 - rather immoral 3 - neutral 4 - rather moral 5 - very moral.

The three persons of the committee will submit an assessment for your decision as well as the decision of your partner.

How do you receive your payment?

After the committee has assessed the decisions, the committee will give you the assessments of both your decision and the decision of your partner, and the person responsible for the payments will give you your payment. In the event that you have decided to donate, you will receive a donation confirmation.

Data protection

The subsequent analysis of all data is carried out anonymously so that your decision can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decision is only used for anonymized data analysis.

Please note:

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to experimente@briq-institute.org after the study.

B.5 Treatment *MPL Low Image*

B.5.1 Your Decision

You will soon have the possibility to choose in 21 decision scenarios between two options: option A and option B. Both options are as follows:

Option A

Option A: I save a human life. By choosing option A, you save a human life. Specifically, by choosing option A, you instigate a donation of 350,00€ that will ensure that at least one person is saved from death by Tuberculosis, just as described before. If you choose option A, you will not receive an additional payment.

Option B

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment. In addition, the absence of your donation will cause the death of a human life.

Additional Payment

The additional payment that you receive from choosing option B varies in each of the 21 decision scenarios. In the first scenario, the payment is 0€ and then increases incrementally in each scenario thereafter by 10€ up to a payment of 200€. Therefore, the decision scenarios look as follows:

Automatic Completion Help

So that you do not need to click as much, we have activated an automatic completion help that automatically fills out the fields for you. As soon as you choose an amount from option B, we assume that you would choose all respectively higher payments from option B. Likewise, when you choose option A in a row, we assume that you would choose option A over all respectively lower payments from option B.

Please note: You can always change your decisions until you clicked on "Confirm Decisions". Therefore, only click on that button when you are certain how you want to decide.

Payment

After you have selected one of the two options for each of the 21 decision scenarios, one of them will be randomly selected for real implementation. This means that the consequences of this decision will be implemented exactly as stated. Each of the 21 scenarios has the same probability of being selected. Therefore, since each of your decisions is potentially relevant, it is in your interest to decide in every scenario as if that decision is being implemented for real.

Your partner sees exactly the same 21 decision scenarios as you and, like you, makes a decision for every scenario. Furthermore, for you and your partner, the same decision scenario will be randomly selected. Thus, both your decision and the decision of your partner for this scenario will be implemented.

The following examples elaborate on this. Assume that decision scenario 2 is randomly selected, and you chose option A, while your partner chose option B. Then you save a human life and your partner will receive 10€. If, on the contrary, both of you choose option B, then both of you will receive 10€. If both of you choose option A, then two human lives will be saved. Assuming that decision scenario 21 is randomly selected, and you chose option B, while your partner chose option A. Then, you will receive 200€ and your partner saves a human life. If, however, both of you chose option B, then both of you will receive 200€. If both of you chose option A, then two human lives will be saved.

Summary

On the page after next, you will make a decision for 21 scenarios, and in each decision, you can choose between option A and option B. By choosing option A, you save a human life, whereas by choosing option B, you receive an additional payment. After you have reached all of your decisions, one of the 21 scenarios will be chosen randomly for you and your assigned partner. Thereafter, the consequences of the chosen decision are realized, i.e., in the case that you chose option A under this scenario, a donation will be made towards the saving of a human life and in the case that you chose option B, you receive the respective amount from the selected scenario. The same applies to your partner. On the next page, you will receive details about the payment procedure.

B.5.2 Further Procedure

After you confirmed your decisions on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. Thereafter, some additional questions will follow. After you have answered these questions, you will get the screenshot with the decisions of your partner displayed, and your partner will get the screenshot with your decisions. You will not receive any further information about your partner, and your partner will not receive any further information about you.

After you received the screenshot, please remain seated until you are called with your cabin number. Then you can go into the adjacent room to pick up your compensation for today's experiment. You will be called one by one so that there is no contact with other participants of the experiment.

Who will be in the adjacent room?

In the adjacent room, you will find the participant who was randomly selected from all participants at the start of the study.

How do you receive your payment?

This participant will give you a sealed envelope with your payment. The selected participant has already received the envelope sealed. Since this participant is only responsible for the payment, this participant has not completed the study and therefore has no knowledge of the decisions to be made. Therefore, this participant does not know what you chose, how you decided, or how much money you received, exactly as explained at the beginning of the study. By handing in your note with your cabin number, you will receive the envelope intended for you.

Data protection

The subsequent analysis of all data is carried out anonymously so that your decisions can never be linked to your person. Your anonymity is therefore always guaranteed, and the

information about your decisions is only used for anonymized data analysis.

Please note:

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to experimente@briq-institute.org after the study.

B.6 Treatment *MPL High Image*

B.6.1 Your Decision

You will soon have the possibility to choose in 21 decision scenarios between two options: option A and option B. Both options are as follows:

Option A

Option A: I save a human life. By choosing option A, you save a human life. Specifically, by choosing option A, you instigate a donation of 350,00€ that will ensure that at least one person is saved from death by Tuberculosis, just as described before. If you choose option A, you will not receive an additional payment.

Option B

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment. In addition, the absence of your donation will cause the death of a human life.

Additional Payment

The additional payment that you receive from choosing option B varies in each of the 21 decision scenarios. In the first scenario, the payment is 0€ and then increases incrementally in each scenario thereafter by 10€ up to a payment of 200€. Therefore, the decision scenarios look as follows:

Automatic Completion Help

So that you do not need to click as much, we have activated an automatic completion help that automatically fills out the fields for you. As soon as you choose an amount from option B, we assume that you would choose all respectively higher payments from option B. Likewise, when you choose option A in a row, we assume that you would choose option A over all respectively lower payments from option B.

Please note: You can always change your decisions until you clicked on “Confirm Decisions”. Therefore, only click on that button when you are certain how you want to decide.

Payment

After you have selected one of the two options for each of the 21 decision scenarios, one of them will be randomly selected for real implementation. This means that the consequences of this decision will be implemented exactly as stated. Each of the 21 scenarios has the same probability of being selected. Therefore, since each of your decisions is potentially relevant, it is in your interest to decide in every scenario as if that decision is being implemented for real.

Your partner sees exactly the same 21 decision scenarios as you and, like you, makes a decision for every scenario. Furthermore, for you and your partner, the same decision scenario will be randomly selected. Thus, both your decision and the decision of your partner for this scenario will be implemented.

The following examples elaborate on this. Assume that decision scenario 2 is randomly selected, and you chose option A, while your partner chose option B. Then you save a human life and your partner will receive 10€. If, on the contrary, both of you choose option B, then both of you will receive 10€. If both of you choose option A, then two human lives will be saved. Assuming that decision scenario 21 is randomly selected, and you chose option B, while your partner chose option A. Then, you will receive 200€ and your partner saves a human life. If, however, both of you chose option B, then both of you will receive 200€. If both of you chose option A, then two human lives will be saved.

Summary

On the page after next, you will make a decision for 21 scenarios, and in each decision, you can choose between option A and option B. By choosing option A, you save a human life, whereas by choosing option B, you receive an additional payment. After you have reached all of your decisions, one of the 21 scenarios will be chosen randomly for you and your assigned partner. Thereafter, the consequences of the chosen decision are realized, i.e., in the case that you chose option A under this scenario, a donation will be made towards the saving of a human life and in the case that you chose option B, you receive the respective amount from the selected scenario. The same applies to your partner. On the next page, you will receive details about the payment procedure.

B.6.2 Further Procedure

After you confirmed your decisions on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. Thereafter, some additional questions will follow. After you have answered

these questions, you will get the screenshot with the decisions of your partner displayed, and your partner will get the screenshot with your decisions. You will not receive any further information about your partner, and your partner will not receive any further information about you.

After you received the screenshot, please remain seated until you are called with your cabin number. Then you can go into the adjacent room to pick up your compensation for today's experiment. You will be called one by one so that there is no contact with other participants of the experiment.

Who will be in the adjacent room?

In the adjacent room, you will find a person who will make the payment. As mentioned before, there will also be a committee consisting of three other persons in the adjacent room. These three persons are students of the University of Bonn and were specially selected for this task.

What is the task of the committee?

The task of these three persons is to assess the decisions you and your partner have taken. Specifically, the assessment is about how moral your behavior and the behavior of your partner is. Apart from the assessment, the three persons will not interact with you (or with your partner) in any way, and the rating will not influence at all the consequences of your decisions or your payment.

What information does the committee receive?

In order for the committee to make the assessment, the two screenshots of the decisions of you and your partner are projected side by side onto the room's wall using a projector, visibly for all persons in the adjacent room. You are identified by (and only by) your cabin number. For better identification, based on your decisions and the decisions of your partner, you must also say the following two sentences aloud. The first sentence refers to your decisions, the second sentence to your partner's decisions.

Sentence 1: "I have decided from a payment of X€ onwards to take the payment for myself instead of saving human life."

Sentence 2: "My partner has decided from a payment of X€ onwards to take the payment for himself instead of saving human life."

The payment X denotes the amount of money for which you switched from option A to option B for the first time. If you have not decided to take the money in any decision-making situation, i.e., have not switched, you have to say the following as the first sentence:

Sentence 1: "I have decided for no amount to take the payment for myself instead of saving human life."

Similarly, if your partner has not decided to take the money in any decision-making situation, you must say the following second sentence:

Sentence 2: “My partner has decided for no amount to take the payment for himself instead of saving human life.”

In summary, you have to say two sentences, and the following information will be visible to everyone in the room:

- The complete list of all 21 decision scenarios described before.
- How you and your partner have chosen in each of these scenarios. This means that for each payment amount, one can see whether you have decided to save a human life or the additional payment and whether your partner has decided to save a human life or the additional payment.

How does the assessment work?

The committee will assess your decisions using a scale. For this, each one of the three persons of the committee selects one of the following five values:

1 - very immoral 2 - rather immoral 3 - neutral 4 - rather moral 5 - very moral.

The three persons of the committee will submit an assessment for your decisions as well as the decisions of your partner.

How do you receive your payment?

After the committee has assessed the decisions, the committee will give you the assessments of both your decisions and the decisions of your partner, and the person responsible for the payments will give you your payment. In the event that you have decided to donate, you will receive a donation confirmation.

Data protection

The subsequent analysis of all data is carried out anonymously so that your decisions can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decisions is only used for anonymized data analysis.

Please note:

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to experimente@briq-institute.org after the study.

B.7 Robustness Experiment

B.8 Introduction

All instructions were displayed on the subjects' screens. The following introduction was the same for both treatments of the robustness experiment.

B.8.1 Welcome to the study

Welcome, and thank you for your interest in today's study!

Please note that you can take part in this study only once. Furthermore, you may only participate if you have registered for this study in our participation database (experimente.bonneconlab.uni-bonn.de).

For your full participation, you will receive a fixed payment of 3€. In this study, you will make decisions on the computer. Depending on how you choose, you can earn additional money. After the study, you will receive all payments, i.e. both the remuneration for your participation and any additional payments based on your decisions, by bank transfer.

On the next screens, you will receive specific information about participation in this study. To proceed, click "Next".

B.8.2 Your Partner

As part of this experiment, a partner has been assigned to you. This partner is a participant in today's experiment, just like you. He or she was randomly assigned to you and will receive the same instructions as you.

In today's experiment, you and your partner will both receive the exact same information and subsequently face the exact same decisions. These decisions have certain consequences, which will be described in detail later.

Payment

At the end of today's experiment, one pair will be randomly drawn from every 24 participants in the experiment. Only the decisions of this pair will be implemented, as described in the instructions. Please note: The random draw of a pair is completely independent of the participants' decisions. Each pair has the same probability of being drawn. Since your decision can be actually implemented for real, you should think carefully about how you will decide in the experiment.

B.8.3 Information

What follows is some information that is relevant to the decisions you will later be asked to make. It concerns the official shop of the University of Bonn.

The Campus Store Uni-Bonn is the official shop of the University of Bonn. Here you can purchase various products such as T-shirts, sweatshirts or mugs with the logo and design of the Uni-Bonn.

The Uni-shop is located at the information point in the main building. There is also an online shop, which can be reached via the website: https://www.campusstore-unibonn.de. The online shop dispatches all goods within 2-3 working days.

Voucher

The next decisions will concern a voucher for the Uni-shop, namely a voucher worth 35€. The voucher can only be redeemed in the online shop and cannot be converted into money.

B.9 Treatment *DE No-Image*

B.9.1 Your Decision

You will soon have the possibility to choose between two options: option A and option B. Both options are as follows:

Option A

Option A: I choose the voucher. By choosing option A, you will receive the voucher for the Uni-shop. Specifically, option A allows you to receive a voucher worth 35€, which you can redeem in the Uni-shop (and only there). If you choose option A, you will not receive an additional payment.

Option B

Option B: I choose 10€ as payment for myself. By choosing option B, you will receive an additional payment of 10€ at the end of the experiment, but you will not receive the voucher.

Additional Payment

Before today's experiment, various amounts between 0€ and 20€ were taken into account for the amount of money you will receive when choosing option B, from which 10€ was selected. Your partner sees exactly the same options as you and makes a decision just like you. So your partner also decides between option A (voucher) and option B (keeping 10€ to himself/herself).

Summary

You will decide on the next page of the screen by choosing between option A and option B. By choosing option A, you receive a voucher. By choosing option B, you receive an additional payment of 10€. On the next page, you will find details about the payment procedure.

B.9.2 Further Procedure

After you confirmed your decision on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. At the end of today's experiment, you will get the screenshot with the decision of your partner displayed, and your partner will get the screenshot with your decision. You will not receive any further information about your partner, and your partner will not receive any further information about you.

Data protection

The subsequent analysis of all data is carried out anonymously so that your decision can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decision is only used for anonymized data analysis.

Please note:

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to experiment@briq-institute.org after the study.

B.10 Treatment *MPL No-Image*

B.10.1 Your Decisions

You will soon have the possibility to choose in 21 decision scenarios between two options: option A and option B. Both options are as follows:

Option A

Option A: I choose the voucher. By choosing option A, you will receive the voucher for the Uni-shop. Specifically, option A allows you to receive a voucher worth 35€, which you can redeem in the Uni-shop (and only there). If you choose option A, you will not receive an additional payment.

Option B

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment, but you will not receive the voucher.

Additional Payment

The additional payment that you receive from choosing option B varies in each of the 21 decision scenarios. In the first scenario, the payment is 0€ and then increases incrementally in each scenario thereafter by 1€, up to a payment of 20€. Therefore, the decision scenarios look as follows:

Automatic Completion Help

So that you do not need to click as much, we have activated an automatic completion help that automatically fills out the fields for you. As soon as you choose an amount from option B, we assume that you would choose all respectively higher payments from option B. Likewise, when you choose option A in a row, we assume that you would choose option A over all respectively lower payments from option B.

Please note: You can always change your decisions until you clicked on “Confirm Decisions”. Therefore, click on that button only when you are certain how you want to decide.

Payment

After you have selected one of the two options for each of the 21 decision scenarios, one of them will be randomly selected for real implementation. This means that the consequences of this decision will be implemented exactly as stated. Each of the 21 scenarios has the same probability of being selected. Therefore, since each of your decisions is potentially relevant, it is in your interest to decide in every scenario as if that decision is being implemented for real.

Your partner sees exactly the same 21 decision scenarios as you and, like you, makes a decision for every scenario. Furthermore, for you and your partner, the same decision scenario will be randomly selected. Thus, both your decision and the decision of your partner for this scenario will be implemented.

The following examples elaborate on this. Assume that decision scenario 2 is randomly selected, and you chose option A, while your partner chose option B. Then you will receive the voucher and your partner will receive 1€. If, on the contrary, both of you chose option B, then both of you will receive 1€. If both of you chose option A, then you and your partner will each receive the voucher. Assuming that decision scenario 21 is randomly selected, and you chose option B while your partner chose option A, then you will receive 20€, and your partner will receive the voucher. If, however, both of you chose option B, then both of you will receive 20€. If both of you chose option A, then you and your partner will each receive the voucher, etc.

Summary

On the page after next, you will make a decision for 21 scenarios, and in each decision, you can choose between option A and option B. By choosing option A, you receive a voucher, whereas by choosing option B, you receive an additional payment. After you have reached all of your decisions, one of the 21 scenarios will be chosen randomly for you and your assigned partner. Thereafter, the consequences of the chosen decision are realized, i.e., in the case that you chose option A under this scenario, you will be given the voucher and in the case that you chose option B, you will receive the respective amount from the selected scenario. The same applies to your partner. On the next page, you will receive details about the payment procedure.

B.10.2 Further Procedure

After you confirmed your decision on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. At the end of today's experiment, you will get the screenshot with the decision of your partner displayed, and your partner will get the screenshot with your decision. You will not receive any further information about your partner, and your partner will not receive any further information about you.

Data protection

The subsequent analysis of all data is carried out anonymously so that your decisions can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decisions is only used for anonymized data analysis.

Please note:

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to experiment@briq-institute.org after the study.