Eliciting Moral Preferences: Theory and Experiment

Roland Bénabou[1], Armin Falk[2], Luca Henkel[3], Jean Tirole[4]

This version: September 20, 2020[5]

[1] Princeton University, NBER, CEPR, briq, IZA, BREAD, and THRED.

[2] Institute on Behavior and Inequality (briq) and University of Bonn.

[3] University of Bonn.

[4] Toulouse School of Economics (TSE) and Institute for Advanced Study in Toulouse (IAST), University of Toulouse Capitole.

[5] We are grateful to Jean-François Bonneton and Marie-Claire Villeval for valuable comments. Ana Luisa Dutra, Juliette Fournier, Pierre-Luc Vautrey, Ben S. Young and Youpeng Zhang provided superb research assistance. 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 Programme Grant Agreement no. 249429 and no. 340950, as well as European Union's Horizon 2020 research and innovation programme, Grant Agreement no. 669217). 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).

Abstract

We examine to what extent a person's moral preferences can be inferred from observing their choices, for instance via experiments, and in particular, how one should interpret certain behaviors that appear deontologically motivated. Comparing the performance of the direct elicitation (DE) and multiple-price list (MPL) mechanisms, we characterize in each case how (social or self) image motives inflate the extent to which agents behave prosocially. More surprisingly, this signaling bias is shown to depend on the elicitation method, both per se and interacted with the level of visibility: it is greater under DE for low reputation concerns, and greater under MPL when they become high enough. We then test the model's predictions in an experiment in which nearly 700 subjects choose between money for themselves and implementing a 350€ donation that will, in expectation, save one human life. Interacting the elicitation method with the decision's level of visibility and salience, we find the key crossing effect predicted by the model. We also show how certain "Kantian" postures (turning down all prices in the offered range) can easily emerge under MPL when reputation becomes important enough.

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

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

"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

How reliably can people's moral preferences be inferred from their choices, including through experimental methods, and how should such estimates be used to inform policy? Relatedly, how should one interpret behaviors that seem deontologically rather than consequentially motivated, such as refusing any tradeoff involving 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, which remains imperfectly understood. We study 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 any image concerns are present, aggregating information from multiple choices decreases prosocial behavior, due to what we term the *discouragement effect.* Conversely, the uncertainty inherent in unconditional pledges increases prosocial behavior, through two effects that lower the effective cost of virtue signaling, and which we term the *cheap-talk* and *cheap-act* effects. Second, these three effects are jointly at work when comparing standard preference-measurement schemes such as direct elicitation (*DE*) and multiple-price lists (*MPL*), leading them to produce systematically different results. The same applies to common public-goods-provision and charitable-contributions mechanisms that share certain key features with them. Third, this behavioral divergence varies not just 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:* The *discouragement effect* dominates when image concerns are low (but positive), 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 (e.g., Bénabou and Tirole 2006, 2011), which we extend to study agents' decisions across a range of settings that differ in their *choice sets,* their *material consequences* (effective cost to self or others) and their *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 very 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's structure, the three key effects at work, and how they lead to the reversal predictions that form the crux of the experiment, we use a simple, concrete example:

Alex and Bob encounter their friend Chris, who works for an organization that provides volunteer tutors to disadvantaged children in surrounding public schools (PS). The work is the same in all the schools (every full Friday afternoon) and so are the students. The only difference is that some schools are more distant than others. PS10 requires only 10 minutes travel, PS30 half-an-hour, PS40 is 40 minutes away, and PS60 a full hour. In the population of potential tutors everyone has the same opportunity cost of time, but there are two types of social preferences, represented respectively by Alex, who would be willing to travel up to 45 minutes to tutor, and Bob, with a maximum willingness of 15 minutes' commuting.

Chris is in charge of recruiting tutors for PS30. Alex immediately accepts. Bob's first thought is to decline, but this would reveal him as less caring than Alex, to Chris and their whole circle of friends she talks to. For a relatively small (net) cost of 15 minutes per week he can instead look just as good, and will thus accept as long as he cares somewhat about his image; for concreteness, let us say it is worth 20 minutes to him. If instead of Chris, the two friends had encountered Casey, who is in charge of finding tutors for PS40, Alex would again have accepted but Bob would have declined, unwilling to pay a net cost of 25 for a reputational value of 20. In this "separate questions" setting, which corresponds to direct elicitation in experiments, even minor reputational concerns generate pooling that increases prosocial behavior.

Suppose instead 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 enough to distinguish them, making it futile for Bob to volunteer for PS30, since there are no intermediate types. 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 the organization, instead of sending out separate recruiters for each school, has them ask people they approach which ones 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. The richness of options offered generates, as before, a discouragement effect for PS30. Working in the other direction, however, is the fact that accepting all schools up to PS40 (as opposed to only PS10) now has an expected cost of only 80/3-10 =17 minutes for Bob. Therefore, if that is what Alex signs up for, Bob will do so as well. Compared to the previous scheme, this second effect thus increases contributions at both 30 and 40. If, moreover, Alex's value of separating from Bob is worth at least 140/4 - 80/3 = 8 minutes of commuting, Alex will sign up for all schools, including PS60; but then Bob will follow suit, and both friends will end up volunteering "unconditionally". 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). The stronger are individuals' reputational concerns, moreover, the stronger is 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 a complementarity.[1]

The standard (Becker-DeGroot-Marschak) experimental implementation of multiple price lists brings in another but related *"cheap talk"* effect, which is the possibility of not being called upon to contribute at all. Such is the case if, for every volunteer, the organization will draw at random (e.g., get a phone call from) any of the four schools. If it is in the volunteer's "acceptable" list on the form 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

[1] This holds for the uniform distribution, and for all other distributions with a monotone hazard rate.

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 accepting all four schools on the list will be called upon to tutor for sure, just as when acceding to a single-school request. While important in general (at intermediate values), this "cheap talk" effect thus plays no role when image concerns are important enough. 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 main 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 (based on detailed medical evidence) in the expected *saving of one human life*; (ii) take money for themselves, where the amount is either a fixed 100€ under *DE* (simple money vs. donation decision), 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. In particular, 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 predicted by the theory. In the *Low Image* treatments, the fraction opting to save a life is 48% under MPL, versus 59% under DE*;* in the *High Image* condition it is 63% under DE, versus 72% under *MPL.* As to subjects who genuinely treat a human life as "priceless," it is of course impossible to identify their number without an infinite budget, but our experiment allows us to bound it from above (probably quite generously) by 26%, which is the fraction who turn down 200€ in order to save a life under the low-image *MPL* treatment.

The paper's results provide a methodological caveat for experiments and contingent-valuation surveys, as well as potential guidance for maximizing public-goods contributions. They also fit well with, and can help account for, the common propensity to respond to moral dilemmas through "righteous" indignation and postures rather than cost-benefit analysis.

1.1 Related Literature

A substantial literature, both theoretical and experimental, studies the interactions between moral behavior, material or social incentives, and markets. Within it, the paper connects together two largely separate lines of work. The first centers on the role of *signaling or identity concerns* (e.g., Bénabou and Tirole 2006, 2011a,b, Ellingsen and Johannesson 2008, Mazar et al. 2008, Ariely et al. 2009, Ashraf et. al. 2012, Lacetera et al. 2012, Exley 2016, DellaVigna et al. 2016, Gino et al. 2016, Grossman and van der Weele 2017, Bursztyn et al. 2018, Falk 2019). 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, Falk and Szech 2013, Elias et al. 2016, Ambuehl 2017, Falk et al. 2020). In analyzing the interactions

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

between the two, the paper also relates to recent 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 Truyts 2019, Bos and Pollrich 2020).

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 (e.g., the strategy-method) affects behavior in one-shot, anonymous games such as dictator, ultimatum, trust or public-goods (e.g., Brandts and Charness 2011, Chen and Schonger 2016a).[3] There is also a large a body of research on elicitation methods for risk and time preferences (e.g., Charness et al. 2013 , Cox et al . 2015, Cohen et al. 2020). There is, to our knowledge, none for reputationally sensitive decisions like those analyzed here.

As discussed earlier, the paper also relates to the long-standing debate in moral philosophy between consequentialist and deontological ethics. Utilitarians justify normative principles and actions by their consequences, aiming at maximizing the Good –pleasure, happiness, desire satisfaction, or welfare (Bentham, 1789, Mill 1861). In contrast, Kantian rule-based ethics value actions *per se,* thus giving priority to the Right over the Good. This means, essentially, a *prohibition on tradeoffs* between personal or even social gains and any harm to others, or to (future) self. Such reasoning is commonly invoked to reject "repugnant" transactions and argue that certain fundamental human (or divine) values are incommensurable with money. As to how people behave *in practice,* the literature on cooperation and voluntary contribution to public goods provides evidence that choices are generally sensitive to the implied consequences for others (Kagel and Roth 1995, Goeree et al. 2002). Likewise, 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 also evidence of "warm glow" altruism in which utility seems to be derived from the act as such, with donations fairly insensitive to the level of contributions by others or that of government funding (e.g., Andreoni 1989, 1990, DellaVigna et al. 2019).

Experiments in which subjects choose between money and a charitable act under varying probabilities that their decision will actually be implemented also point to a mix of motives. Feddersen et al. (2009) and Chen and Schonger (2016b) note that if the same probability applies to costs and benefits, the behaviors of pure consequentialists (in the traditional sense that excludes image concerns) and that of people with purely expressive or/and deontological preference should both be invariant to the odds. In practice, they find that the implementation probability does matter, which reveals a tradeoff between the two types of motives. In Falk, Neuber and Szech (2020), participants in a group decide simultaneously on whether to generate a negative externality in return for money. The frequency of doing so is found to increase as the probability of being pivotal in the outcome falls, in line with consequentialism. However, even as the probability reaches zero there remain about 18% of subjects who turn down the money "on principle," even though they understand (as verified by elicited beliefs) that the chance this will actually do any good is zero.

[3] Concerning DE with deterministic versus random implementation (which can be considered 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 aversion plays no role in the effects we identify, which directly affect expected returns.

2 Model

1. Preferences. We start here from the basic framework in Bénabou and Tirole (2006, 2011), which we will extend to richer choice situations such as multiple-price elicitation mechanisms. Agents are risk-neutral and have a two-period horizon, $t = 1, 2$. At date 1, an individual has an opportunity to engage in moral behavior ($a = 1$) or act selfishly ($a = 0$). Choosing $a = 1$ is prosocial in that it involves a personal cost $c > 0$ but generates a valuable externality or public good, normalized to $e \in [0, 1]$; for instance, e may be the probability of an externality of fixed size 1. Agents differ by their intrinsic motivation to act morally: given e, it is either $v_H e$ (high, moral type) or $v_L e$ (low, immoral type), with probabilities ρ and $1 - \rho$ and $v_H > v_L \geq 0$; the average type will be denoted as $\bar{v} = \rho v_H + (1 - \rho)v_L$.

Besides the externality, the second feature of action $a = 1$ that ties it to the moral domain is that it can be reputationally valuable, conferring on the individual a social or self-image benefit at date 2. In the social context, the agent knows his private type but the audience (peer group, firms, potential partners) does not. In the self-signaling context, the individual has an immediate, "intuitive" sense of his deep preferences at the moment of action –say, how much empathy or spite he experiences– but later on the intensity of that feeling is imperfectly accessible ("forgotten"); 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 choosing action $a \in \{0, 1\}$ and μ measures the strength of self or social-image concerns, common to all agents. This utility level may be additively augmented by any externalities or public goods generated by others, but since this term is independent of the agents' action we omit it here.

Note that these preferences are consequentialist: an agent's desire to behave prosocially reflects and trades off the externality he expects his actions to have, the personal costs involved, and the reputational consequences.[4]

2. Behavior. As is common in signaling models, multiple equilibria may coexist. For instance, 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. Intuitively, when the high type is expected to abstain the stigma from doing so is lessened, which in turn reduces the net reputational value of acting morally. In this and all other cases of multiplicity (see Appendix A), we choose the equilibrium that is best for both types, namely the no-contribution pooling equilibrium. Indeed, the separating equilibrium yields lower payoffs both for the low type, $\mu \cdot v_L < \mu \bar{v}$, and

[4] The model also allows for genuinely deontologica" agents, $v_H = +\infty$, but a key point is that they are not required to generate "observationally deontological" behavior in MPL-like experiments, i.e., subjects rejecting all prices (for harming someone else) on some randomly implemented price list with finite or even infinite support.

for the high one, $v_H e - c + \mu v_H \leq \mu \bar{v}$.

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. As discussed in Bénabou et al. (2018), these basic predictions align quite well with a broad range of empirical evidence.

3 Measuring Moral Preferences

We now expand the model to study what can be learned from people's willingness to accept different tradeoffs between personal gain and harm for others, or their "Kantian-like" refusal to accept any among those proposed. Two main types of situations are considered, corresponding respectively to the *DE* and *MPL* preference-elicitation methods. In the first one, the individual is faced with a take-it-or-leave-it opportunity to incur a given cost (or forfeit a given prize) c, in order to create an external benefit. In the second, more auction-like context, he can "name his price" from within a given range, and this is made incentive compatible by randomly drawing some cost $\tilde{c}$ within it, and making him take the action at cost $\tilde{c}$ only when it is below his stated willingness to pay, c.

3.1 Direct Elicitation

We first derive an agent's choice of how to act when faced with any *given* value $c \in \mathbb{R}_+$ of the cost of moral action. The strength of reputational concerns μ can also be modulated, by varying the public visibility or private memorability of choices. Equilibrium behavior will be fully characterized by three cost (or incentive) thresholds, strictly increasing in μ and corresponding to the three straight lines in Figure 1:

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

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

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

Denoting $a_H^{DE}(c,\mu)$ and $a_L^{DE}(c,\mu)$, or a_H and a_L for short, the strategies (probabilities of choosing $a = 1$) of the two types, we show the following intuitive results, illustrated in Figure 1 (for the case $\rho < 1/2$).

Proposition 1 (direct elicitation). *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 still behaves morally ($a_H = 1$), but the low type's probability $a_L(c)$ of doing so decreases with c, and is then equal to 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" (i.e., 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

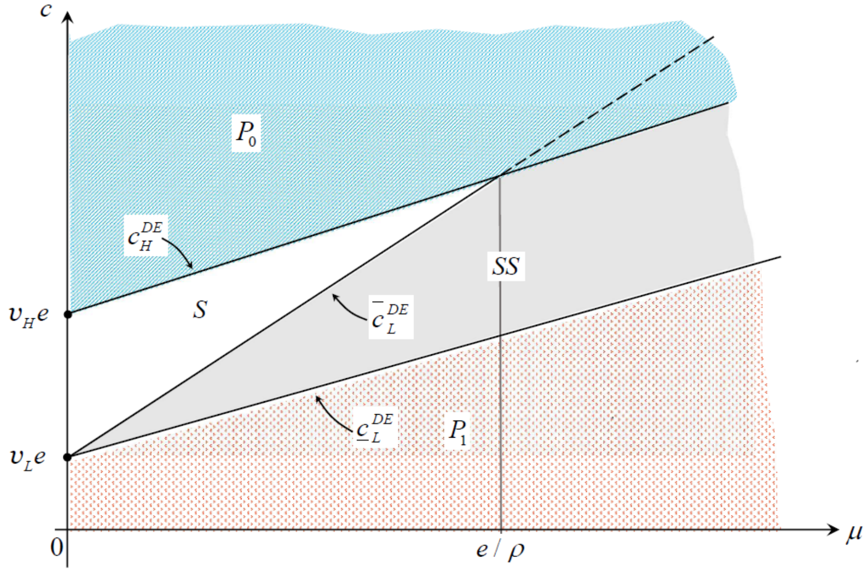

Figure 1: Direct Elicitation. P_0: pooling at $a_H = a_L = 0$; S: separation, $a_H = 1, a_L = 0$; SS: semi-separation: $a_H = 1, a_L \in (0,1)$; P_1: pooling at $a_H = a_L = 1$.

separation (and hence full revelation of actual preferences) occurs shrinks with μ, and becomes empty for $\mu > e/\rho$.

3.2 Multiple-Price List

We now turn to a standard *MPL* or Becker-DeGroot-Marschak-type mechanism. Each subject is asked what maximum cost c he is willing to incur for taking the moral action, $a = 1$, knowing that the actual $\tilde{c}$ will be drawn according to some cumulative distribution $G(\tilde{c})$ on an interval $[0, c_{\max})$, with $0 < v_L e < v_H e < c_{\max}$. Equivalently, this cutoff value represents his willingness to accept a "bribe" to make the immoral choice, $a = 0$. In experiments the distribution G is typically uniform, but here we allow other cases as well, including $c_{\max} = +\infty$. Let $L(c)$ denote the low type's net loss associated with refusing any reward below level 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 $E_G[\tilde{c}]$ be finite. We will say that a subject is *observationally deontological* if he turns down all prices on the proposed list (with distribution G): he behaves, given the available data, as someone who would not act immorally "at any price."

We now solve for both types' willingnesses 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; later on, we will compare them to their counterparts under direct elicitation. Note first that, *absent reputation concerns,* both mechanisms are *equivalent* and reveal each type's *true* moral preference: $c_H^{DE} = c_H^{MPL} = v_H e$, $\underline{c}_L^{DE} = \bar{c}_L^{DE} = c_L^{MPL} = v_L e$. When $\mu > 0$, there are again three cases to consider, defined

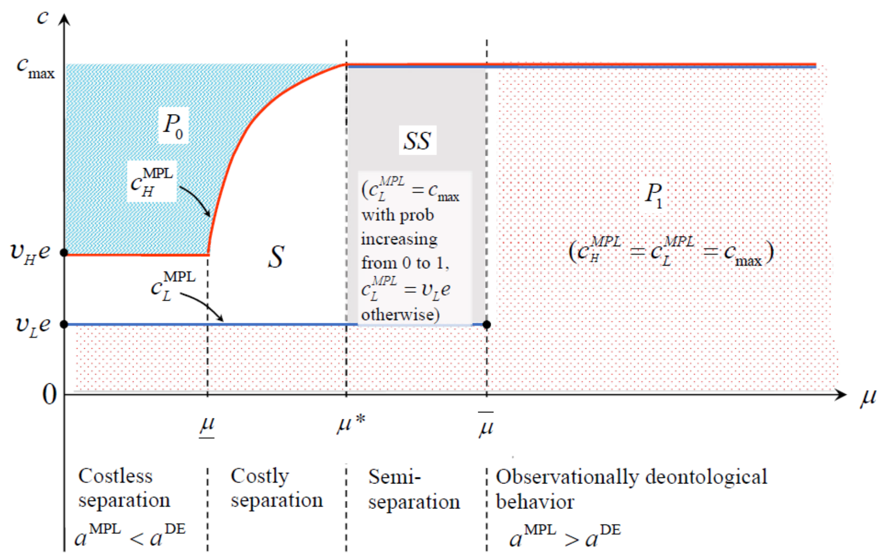

Figure 2: Multiple-Price List. P_0: pooling at $a_H = a_L = 0$; S: separation, $a_H = 1, a_L = 0$; SS :semi-separation: $a_H = 1, a_L \in (0, 1)$;P_1: pooling at $a_H = a_L = 1$.

by the thresholds

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

and illustrated in Figure 2. \

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

1. Separation: *when the (self) reputational concern μ is low, $\mu < \mu^*$, the high type's WTA for behaving immorally is $c_H^{MPL} = \max\left\{v_H e,\ L^{-1}(\mu(v_H - v_L))\right\}$, 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. Semi-separation: *when μ is intermediate, $\mu \in [\mu^*, \bar{\mu}]$, the low type seeks to pool while the high type seeks to separate, so their WTA's escalate until the high type becomes observationally deontological, $c_H^{MPL} = c_{\max}$, while the low type randomizes between that same "virtuousness" ($c_L^{MPL} = c_{\max}$) and revealing himself (accepting $c_L^{MPL} = v_L e$), with a probability $a_L(\mu)$ increasing in μ.*

3. Pooling: *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, the presence of image concerns naturally raises contributions, as can be previewed from Figures 1-2. 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 total contributions?*

Each of these can be asked in two forms:

(a) At a given cost $c \in [0, c_{\max}]$: what fraction of people $\bar{a}^{DE}(c,\mu)$ accept to pay or forfeit c in order 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*, thereby committing to pay c and implement $a = 1$ if c turns out to be the randomly drawn cost? And how does the sign of $\bar{a}^{DE}(c,\mu) - \bar{a}^{MPL}(c,\mu)$ depend on μ? This will correspond exactly to what our experiment does.

(b) Aggregating over costs $G(c)$: suppose an individual or population is confronted with a distribution of costs $G(c)$, one by one under *DE*, or via a single list under *MPL*. How do the resulting contributions,

$$\bar{a}^{DE}(\mu) = \rho G(c_H^{DE}) + (1-\rho)\int_0^{\min\{\bar{c}_L^{DE}, c_H^{DE}\}} a_L(\tilde{c})dG(\tilde{c}), \tag{6}$$

$$\bar{a}^{MPL}(\mu) = \rho G(c_H^{MPL}) + (1-\rho)G(c_L^{MPL}), \tag{7}$$

compare, and does the answer depend on μ?

While the answers generally depend on the specific value of c, or distribution $G(c)$, the cases of low enough and high enough image concerns yield clear predictions. In stating them, it will be useful to denote as μ^{**} the solution to $\underline{c}_L^{DE}(\mu) = c_{\max}$, or

$$\mu^{**} \equiv \frac{c_{\max} - v_L e}{\bar{v} - v_L} > \frac{L(c_{\max})}{\bar{v} - v_L} = \bar{\mu}. \tag{8}$$

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 $c \in (v_L e, \underline{c}_L^{DE}(\mu))$ and $c \in (v_H e, \bar{c}_H^{DE}(\mu))$, both nonempty.*

3. *For all $\mu \geq \bar{\mu}$, $a_\tau^{DE}(c,\mu) \leq a_\tau^{MPL}(c,\mu)$, 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 intuitive effects.

1. Weak image concerns: discouragement effect dominates. When $\mu > 0$ is low enough that separation is costless under *MPL,* 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 first result. Intuitively, *MPL* raises the cost to the low type of mimicking the high one, since to do so he must be willing to forego up to $v_H e$, and for low reputational gain (small μ) 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 initially flat in Figure 2 until μ becomes high enough, whereas it is linear in Figure 1.

2. Strong image concerns: cheap-act effect dominates. Conversely, at high values of μ reputational concerns become paramount, and the cost of signaling is lower (for both types) under *MPL* than under *DE,* as high values of c must be paid only with probability less than 1: the effective cost of stating a cutoff c is only $E[\tilde{c} \leq c] < c$. 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 $\mu < \mu^{**}$ (meaning that the low type does not pool at $c_{\max}$). Note, furthermore, that for any distribution satisfying the monotone likelihood ratio property ($g/(1-G)$ increasing), the function $c - E[\tilde{c} \leq c]$ is increasing in c. Therefore, as μ rises, and with it each type's *MPL* cutoff, the cheap-act effect becomes stronger, implying a larger contributions gap relative to *DE.*

3. Intermediate image concerns. In the range $(\underline{\mu}, \bar{\mu})$ 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 cheaper (especially for low types) to state high cutoffs, thus reinforcing the cheap-act effect. The latter is not as strong in this range as for high values of μ, and conversely the cheap-signal effect weakens as μ rises, pushing $G(c^{MPL})$ closer to 1. The net balance of the three effects is thus generally ambiguous in this intermediate range, and consequently so are the sign of $a^{DE} - a^{MPL}$ and its variation with μ.

Aggregating outcomes over the cost distribution $G(c)$, Proposition 3 also implies that:

Corollary 1. *Let $G(c)$ have positive density on $[0, c_{\max})$.*

1. *$\bar{a}^{DE}(\mu)$ and $\bar{a}^{MPL}(\mu)$ coincide at $\mu = 0$, then both increase until they reach 1. The increase is strict on $[0, \mu^{**}]$ for $\bar{a}^{DE}(\mu)$, and on $[\underline{\mu}, \bar{\mu}]$ for $\bar{a}^{MPL}(\mu)$.*
2. *For all $\mu \in (0, \underline{\mu})$, $\bar{a}^{DE}(\mu) > \bar{a}^{MPL}(\mu)$.*
3. *For all μ $(\bar{\mu}, \mu^{**})$, $\bar{a}^{DE}(\mu) < \bar{a}^{MPL}(\mu)$.*

For the same reasons as above, in the intermediate-reputation range $(\underline{\mu}, \bar{\mu})$, there is generally no unambiguous relationship between the value of μ and the sign of $\bar{a}^{DE} - \bar{a}^{MPL}$. For the uniform distribution universally used in experimental work, however, we are able to provide relatively simple conditions (given in Appendix A) ensuring *a single-crossing property,* namely that $\bar{a}^{MPL}(\mu) > \bar{a}^{DE}(\mu)$ *if and only if* μ exceeds a critical cutoff.

Testable Implications. Three main hypotheses thus emerge from the model, with the last two being most important. First, as usual, greater visibility increases contributions. Second, at low but positive levels of visibility, the *DE* mechanism 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: the *MPL* mechanism induces more moral decisions, due to the now dominating *cheap-act* effect.

The corresponding inequalities in Proposition 3 can be weak or strong, depending on the specific region of the parameter space; this is a standard feature of models with discrete types and action spaces, which typically disappears when there is enough heterogeneity (e.g., a continuum) to span all cases. Corollary 1 provides such a demonstration with cost heterogeneity, and in real subject populations there will be much more heterogeneity in values v than in our simple binary model, resulting in similar smoothing. For this reason, when confronting the model to data we will tighten the predicted inequalities to be strict ones, which of course is more demanding.

4 Experimental Design

We now present an experiment designed to test the model's main predictions. In particular, we will test whether sensitivity to image concerns varies with the mode of elicitation and the strength of image concerns, generating the key *crossing property* described in Proposition 3. For this purpose, we need a design that confronts subjects with a significant moral decision, varying both the elicitation procedure and the importance of social image.

4.1 Saving a Life Paradigm

We first need to generate choice situations in which one option is unambiguously perceived as the more moral one. To that effect, we adopt the *Saving a Life* paradigm, also used by Falk and Graeber (2020), in which a subject can either take money for him or herself 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 top 10 causes of death worldwide.[5] At the same time, there are highly effective treatments available using antibiotics, for which the success rate well exceeds 85%. Together with the Indian non-profit organization *Operation ASHA*, we calculated a specific monetary amount sufficient to identify, treat, and cure a number of patients such that, *in expectation, one person will be saved* from death by tuberculosis *because of 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 (Straetemans et al. 2011, Tiemersma et al. 2011, Kolappan et al. 2008), we determined that level to be 350€ (about $385 at the time of the experiment). By allowing for the treatment of five patients, such a donation allows the (expected) saving of one human life; these are, in fact, conservative estimates.[6]

This paradigm maps directly into our model: contrasting 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) induces a clear tradeoff between morality and self-interest. This is reenforced by the donation being cost-effective in both absolute and relative terms: the amount is well above all monetary payments possible for the subjects themselves, as described later, and

[5] World Health Organization Tuberculosis Fact Sheet (https://www.who.int/en/news-room/fact-sheets/detail/tuberculosis).

[6] When the person whose life or suffering is at stake is known, directly visible (e.g., media attention) or otherwise made distinct even in minor ways to a decision-maker, the latter's willingness to pay rises considerably. This "identifiable victim effect" (Small, Loewenstein and Slovic 2003; Small and Loewenstein 2007) is not a feature of our experiment. While charitable campaigns often leverage its emotional power to (presumably) good ends, it is generally not considered a reliable guide for rational public decisions.

the money is directly used to treat patients, without any administrative or transaction cost.[7]

4.2 Treatments

In the *Saving a Life* paradigm, subjects face the decision to either engage in moral behavior by triggering the donation ($a = 1$) or act selfishly by taking money for themselves ($a = 0$). Our model predicts that the likelihood of moral behavior depends both on the elicitation method and the level of image concerns. Therefore, we used 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 direct elicitation (DE), subjects faced the binary choice between receiving $c = 100€$ as payment, or saving a human life in expectation; Figure 6 in Appendix C shows the decision screen. We predetermined, as part of the experimental design, this single value of $c = 100€$ as a compromise between two practical concerns: (i) c must be high enough to be meaningful (generate choices of both types); (ii) in contrast to *MPL*, each implemented decision has a sure cost to the experimental budget of either c of the full 350€ donation, which given our large subject sample quickly adds up. The latter consideration is, in particular, why we decided to implement *DE* with a single level of c.

In contrast, for the MPL conditions, we used a price-list design: starting with $c = 0€$ and going in 10€ increments up to $c = 200€$ ($\approx$\$220), subjects could indicate in each of the 21 contingent choices whether they want to save a life or take the amount c for themselves (Figure 7 in Appendix C shows the decision screen). Each price was then equally likely to be drawn for implementation (uniform G).[8] The price-list procedure elicits the minimal price c for which a subject prefers the immoral action over the moral one.

Turning now 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, $\mu > 0$; the former is then predicted to dominate at low values, leading to $\bar{a}^{DE}(\mu) > \bar{a}^{MPL}(\mu)$, and the latter at high values, reversing the inequality. To ensure such a minimal level of (self- or social) image concerns in both treatments, we notified subjects at the start that they were anonymously paired with another subject of the same session. They learned that at the end of the experiment, they would be able to see their partner's choices displayed alongside their own decisions on their screens, and that their partner would see the same. Apart from having the opportunity to observe the partner's choices, subjects received no further information about him or her, so that no aspect of the dyad other than the visibility of one's choices to one's partner would influence decisions.

Social-image image concerns can still arise, however, in experiments where the final payment procedure is not anonymous. In such cases, the experimenter can at least partly infer choices from the amounts paid, and link them to subjects' identity when interacting with them; due to the binary nature of the decisions in our experiment, choices are in fact perfectly revealed through the payments. In order to keep the level of image concerns at a minimum in the *Low*

[7] This was made possible through the collaboration of the charity involved.

[8] To avoid inconsistent answers, we enforced a single-switching rule: if a subject chose to save a life for a given monetary value, the computer filled in the same choice for all lower values. However, participants could always revise their decisions after the autocompletion, and had to confirm their final choices before moving on.

Image treatment ($\mu = \mu_L$), we therefore designed the payment procedure to be *double-blind*, so that not even the experimenter could link subject's decisions to their identity. Following Barmettler, Fehr, and Zehnder (2012), subjects received their payment not from the experimenter, but from another subject. At the start of the experiment, one subject was randomly designated to carry out all payments. He or she did not participate in the regular experiment and, as such, 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 the experiment, in an adjacent room.[9]

The *High Image* treatment ($\mu = \mu_H$), in contrast, was designed to induce strong image concerns. Subjects were informed that, upon receiving their payment, their choice would be observed and compared to those of their matched partner by a *committee consisting of three persons,* sitting at a table in the room where payments would take place. Two features ensured that the committee could accurately observe the choices of each subject and their matched partner. Both choices were projected onto a wall, and subjects had to state their choices and those of their partners aloud, using two predetermined sentences. Only afterward did they receive their payments. To provide a justification for the existence of the committee, and further emphasize the focus on moral judgment, subjects were informed that the committee would judge both partners' decision(s) based on their degree of morality. For this, each committee member used a scale with five options, ranging from very immoral to very moral, to rate each subject and their matched partner choice(s). Apart from observing and filling out the scale, the committee did not interact with the subjects in any way.

Compared to other studies, our image intervention features an explicit comparison of choices and a particular emphasis on their moral dimension, through the judgment of the committee. Both were intended to further magnify the extent of image concerns already generated by the increase in the visibility of choices. While quantifying the level of image concerns inherent in a situation is difficult, these features of our image manipulation seem well suited to induce strong image concerns.

4.3 Procedure

In total, 697 subjects (405 female, mean age = 24.01, SD = 6.21) participated in 36 sessions at the BonnEconLab of the University of Bonn. 178 took part 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 ORSEE (Greiner 2004), and the experiment was conducted using oTree (Chen, Schonger, and Wickens 2016). The sessions lasted about 60 minutes on average, and subjects received a show-up fee of 12€ (≈\$13). For each session, one matched pair of subjects was randomly drawn, and their choices were implemented for real. As a consequence, in the DE treatments, each of the two subjects drawn either received the 100€ for themselves, or triggered a 350€ donation to the charity. In the MPL treatments, one price from the price list was randomly drawn (with equal probabilities), and the pre-stated choices of the two subjects for this price were implemented. Therefore, subjects either triggered the donation or received up to 200€.

[9] To make clear that the person responsible for the payment procedure was a randomly determined participant, the experimenter collected from each subject a token indicating their respective cabin number. All tokens were thrown into a bag, and one was drawn in front of all participants.

At the beginning of each session, subjects received a brief verbal introduction to the experiment. In the *Low Image* treatments, the procedure ensuring anonymity was briefly explained and demonstrated. In the *High Image* treatments, the existence of the committee was announced, and subjects walked through a room where a table with three seats and a projector had been set up. Subsequently, subjects received detailed information about tuberculosis, its effects, and treatment; the instructions also referred them to a website where they were invited to confirm the validity of the information. We further introduced the charity and its working procedure to subjects, and explained our calculations regarding the specific donation amount of 350€. Subjects then received information about their choice options and, after answering a couple of comprehension questions, made their decisions. While making the decisions, subjects could always click on a button on the decision page to be reminded of the exact link between the term "saving a life" and a donation of 350€.[10] Subsequently, subjects 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

We start by summarizing how the theoretical predictions of Proposition 3 will be tested.

5.1 Hypotheses

Recall that in all treatments, subjects made a choice between saving a life and 100€. In *DE*, this was their only choice, while in *MPL* it was embedded in a price list with support $[0, 200]$, in increments of 10€. The choice was either made in private or disclosed to an evaluating committee, inducing either a low (μ_L) or a high (μ_H) level of image concerns, respectively. Our outcome variable is the fraction $\bar{a}^i(\mu)$ of subjects who chose to save a life, given an elicitation method $i \in \{DE,\ MPL\}$ and a level of image concerns $\mu \in \{\mu_L, \mu_H\}$. For brevity, we will refer to $\bar{a}^i(\mu)$ as "total contributions". Based on Proposition 3, the predictions to be tested are:

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

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

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

Hypothesis 1 is familiar, capturing 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 that of the *cheap-act effect* at μ_H, respectively. Together, they constitute the distinctive crossing prediction discussed earlier.

[10] The translated instructions can be found in Appendix C.

5.2 Experimental Results

We start by examining Hypothesis 1, i.e., the effect of our manipulation of visibility on moral behavior. Under both elicitation methods, increased visibility led to a rise in mean contributions. However, the magnitude of the increase is markedly different between the two. Under *DE*, 58.8% of subjects chose to save a life in the *Low Image* and 62.5% in the *High Image* treatment, which is a relatively small and insignificant increase ($p = 0.51$, two-sided Fisher's exact test) in contributions.[11] Under *MPL*, increased visibility led to a much larger increase in contributions, as can be seen in Figure 3.

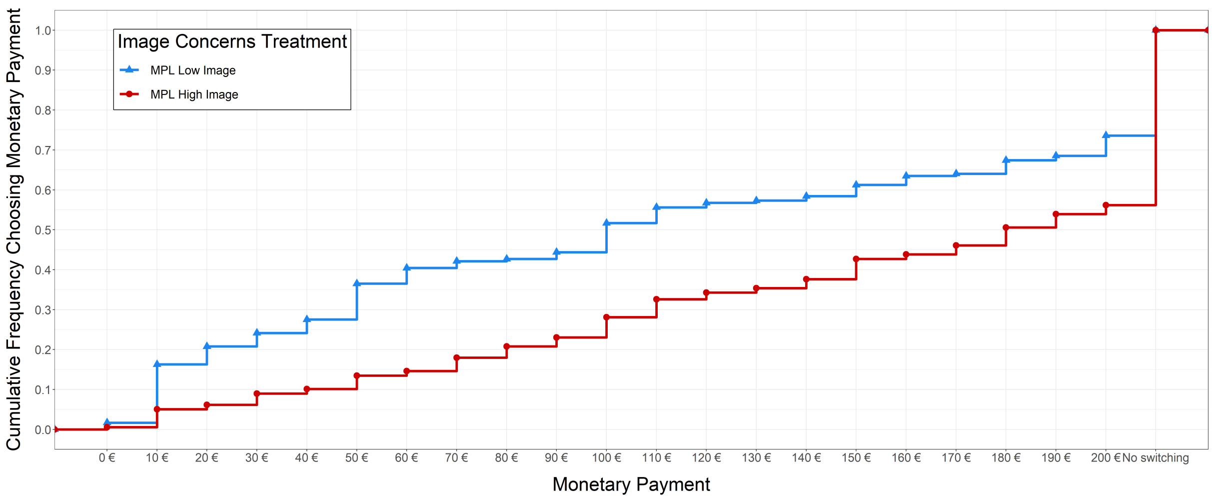

Figure 3: Switching behavior of subjects under MPL. The figure plots the cdf in the Low Image and High Image treatments, for each payment available in the price list. "No switching" indicates the fraction of subjects who refused payment at all available prices.

The figure displays the cumulative distribution function (cdf) of the fraction of subjects choosing the monetary payment, for each $c \in [0, 200]$ offered in the price list. For example, at 40€, about 28% of subjects in the *MPL-Low Image* treatment have already switched from the moral choice to money for themselves. This fraction increases to almost 37% at 50€, meaning that nearly 10% of the subjects switched at this value from saving a life to taking the money. As evident from the figure, the cdf generated by the *Low Image* treatment lies above the cdf of the *MPL-High Image* treatment for all monetary payments ($p < 0.001$, Kolmogorov–Smirnov test). Furthermore, the difference is more than 15% for almost all payments; it is the largest at 60€, with a difference of 26%. In particular, at 100 €, which will be important for the comparison with the *DE* treatments, contributions are 23.6% higher under *High Image* than under *Low Image*, an increase that is significant at any conventional level ($p < 0.0001$, two-sided Fisher's exact test). The differential effects of the image manipulation under *DE* vs. *MPL* are closely linked to the "crossing" pattern predicted by Hypotheses 2 and 3, which we test next.

[11] We follow the convention of reporting two-sided test results, even though we are testing the directed (inequality) hypotheses emanating from the model.

Our main interest is the comparison between the *MPL* and *DE* elicitation methods: Figure 4 displays the fraction of subjects choosing to save a life instead of 100€ for themselves, i.e., total contributions $\bar{a}^i(100, \mu)$, under each of the four treatments. Contributions clearly differ by elicitation method, and the sign of the difference reverses between low and high image, as predicted. In the *Low Image* treatments, we observe $\bar{a}^{MPL}(\mu_L) < \bar{a}^{DE}(\mu_L)$, as predicted by Hypothesis 2, consistent with the dominance of the *discouragement effect* for weak image concerns. The difference is large, with a fraction of subjects choosing to save a life rising from 48.3% to 58.8% between *MPL* and *DE* ($p = 0.065$, two-sided Fisher's exact test). In contrast, in the *High Image* treatments, we observe $\bar{a}^{MPL}(\mu_H) > \bar{a}^{DE}(\mu_H)$, in line with the *cheap-act effect* dominating for strong image concerns, 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 ($p = 0.070$, two-sided Fisher's exact test).

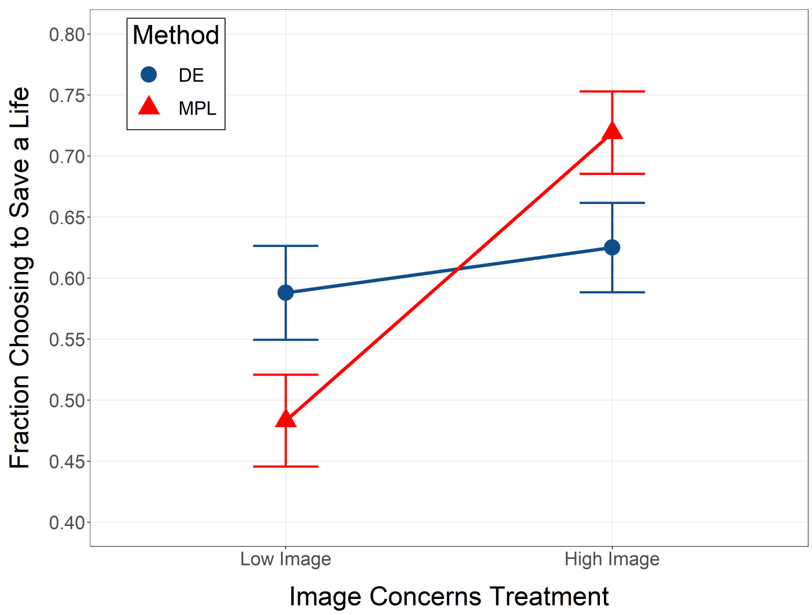

Figure 4: Interaction effect of elicitation methods and image concerns on moral behavior. The figure shows the fractions of subjects choosing to save a life in the MPL and DE treatments, under either the Low Image or the High Image treatment. Error bars indicate the standard error of the mean.

Table 1 provides further evidence using an OLS-regression analysis.[12] Regressing the probability of choosing to save a life (instead of taking 100€) on a dummy for the type of elicitation (1 for *MPL* treatment) yields a positive coefficient for *Low Image* in Column (1), and a negative one for *High Image* in Column (3), respectively. As can be seen in Columns (2) and Column (4) respectively, these effects remain largely unaffected when including a set of control variables that may independently affect moral behavior. These controls include 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 together implies

[12] Our results remain qualitatively unchanged if we use a Probit or Logit regression model.

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

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
R^2	0.011	0.077	0.010	0.062

The table shows OLS regression coefficients. Robust standard errors in parentheses. Controls include age, gender, income, religiousness, educational level, and high school grade. Significance levels: *p<0.1, **p<0.05 and ***p<0.01.

that the increase in contributions as image increases from μ_L to μ_H is more pronounced for *MPL* than for *DE.* To test this, we ran an OLS regression interacting *High Image* with *MPL,* in which *DE - Low Image* is the baseline. As shown in Table 3 (see Appendix B) the interaction is positive and significant at the 1-percent level. In sum, our results support the key predictions (Hypotheses 2-3) emanating from Proposition 3.

5.3 Individual heterogeneity

Note from Figures 1 and 2 that, in both *DE* and *BDM,* it is sometimes the low type v_L who responds more to an increase in visibility μ (trying harder to pool), and sometimes the high type v_H (trying harder to separate), and this again depends on the elicitation, method, as well as the specific range of (c, μ) over which these choices occur. While there are no clear-cut theoretical predictions here, it is nonetheless interesting to examine whether, in our experiment on saving lives, it is low-prosociality or high-prosociality subjects who tend to respond more to publicity, to changes in the elicitation mechanism, or to the interaction of the two.

To that effect, we use a separate survey item on altruism to categorize subjects as either low or high types, respectively. The measure was developed and validated as part of the *Global Preference Survey* (Falk et al. 2018). It consists of the weighted sum of two items, one qualitative and one quantitative. In the qualitative item, subjects are asked how willing they are to give to good causes without expecting anything in return, where answers can be given using a scale from zero to ten. The quantitative item consists of a hypothetical scenario where subjects unexpectedly receive 1,000€ and are asked to state how much of this endowment they would donate.

Table 2: Heterogeneity in the effect of the elicitation method on prosocial behavior

Dependent variable:	Choice to Save a Life (vs. 100€)							
	Below-median Altruism				Above-median Altruism			
	Low Image		*High Image*		*Low Image*		*High Image*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MPL	-0.187^{**}	-0.187^{**}	0.040	0.030	-0.032	-0.007	0.118^{*}	0.138^{**}
	(0.075)	(0.078)	(0.075)	(0.079)	(0.073)	(0.072)	(0.068)	(0.066)
Constant (DE)	0.512^{***}	0.611^{***}	0.592^{***}	0.586^{***}	0.663^{***}	0.591^{***}	0.667^{***}	0.647^{***}
	(0.056)	(0.084)	(0.050)	(0.072)	(0.052)	(0.070)	(0.054)	(0.062)
Controls		X		X		X		X
Observations	168	168	174	174	175	175	180	180
R^2	0.036	0.133	0.002	0.035	0.001	0.101	0.017	0.109

The table shows OLS regression coefficients. Robust standard errors in parentheses. Controls include age, gender, income, religiousness, educational level, and high school grade. Significance levels: *p<0.1, **p<0.05 and ***p<0.01.

Subjects in our experiment were asked these questions as part of a general questionnaire, after making the main donation decision. Reassuringly, choices in the experiment and the altruism measure are significantly correlated (Spearman Rank correlation = 0.21, $p < 0.0001$). More importantly, responses are balanced across treatment conditions; pairwise t-tests reveal that none of the answers are significantly different from each other.[13] To identify subjects as types, we perform a median split by the altruism measure, categorizing below-median subjects as low and above-median subjects as high types, respectively. Table 2 then repeats the regression analysis of Table 1, but separately for low types (Columns (1)-(4)) and high ones (Columns (5)-(8)). We find that under *Low Image*, it is primarily the low-altruism types who contribute less under the *MPL*, consistent with the idea (and the model's intuition) that the discouragement effect is most relevant for this bottom half. For high-altruism types, there is no difference between the elicitation methods. Under *High Image*, this pattern reverses. High-altruism types now contribute significantly more under the *MPL* compared to the *DE*, while there is no difference between the elicitation methods for low-altruism types. These findings suggest that high-altruism types respond more to the cheap-act (and perhaps also, cheap-talk) effect.

6 Discussion and Conclusions

Our analysis yields two main lessons –one concerning potential bias in the measurement of "true" moral preferences, the other related to policies and welfare judgments based on contingent-valuation surveys.

Experiments. Propositions 1-3 show, and our experiment indeed finds, that *image concerns* affect the measurement of moral preferences in ways that *interact with the elicitation method.*

[13] The p-values are $p = 0.71$ for the comparison between *MPL-Low Image* and *DE-Low Image*, $p = 0.16$ when comparing *MPL-High Image* and *DE-High Image*, $p = 0.24$ when testing *MPL-Low Image* versus *MPL-High Image* and lastly $p = 0.95$ for *DE-Low Image* against *DE-High Image*.

This means that regardless of whether one is interested in image-inclusive moral preferences (for *positive* predictions) or in purely intrinsic ones (for *normative* judgements), the estimates will differ between direct and price-list mechanisms. These results contrast with the invariance across methods commonly assumed in experimental work, thus raising the possibility of potential biases in the estimation of moral preferences, and arguing for caution in the interpretation of such estimates.[14]

In particular, even purely utilitarian individuals may act, when facing *MPL*-like situations, as if deontologically motivated: their behavior will resemble that in the initial Kant quote – refusing all proposed prices in exchange for what is perceived as having a dignity. Of course, since only finite (realized and even expected) amounts can be offered, one cannot know how many of them would have accepted, or still refused, even higher prices. 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€ (more than 15 times their show-up fee) in the Low Image *MPL* condition.[15] The latter stems from the fact that this proportion nearly doubles, to 48.35%, with a relatively mild and anonymity-preserving manipulation of the decision's visibility to three anonymous strangers.

A related phenomenon is 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 (limiting access to medical treatments, setting pollution standards or health-and-safety regulations), explicit reference to putting a price tag on life typically produces righteous and conspicuously displayed indignation (e.g., Sandel 2012).

Policy. The second issue is the interpretation and use for policy-making of contingent-valuation surveys. The fact that respondents' stated willingness to pay for non-market goods such as environmental preservation or risk abatement often seems unreasonably high, and invariant to the size of the public good, is usually attributed to the "cheap talk" nature of most such surveys (the actual c faced is zero). Note, however, that even for incentive-compatible elicitations and realistically envisioned tradeoffs (e.g., calibrated from charitable contributions), the willingness to pay includes the value of social and self-image (the latter, even under anonymity). If the purpose of measurement is to predict or explain behavior, this is appropriate (people will actually pay up to this amount), but our results caution that different elicitations methods may yield quite different results. If the purpose is to inform policy, moreover, then even real-money WTP's can substantially overstate the true *social* value of the public good in question. That is because they incorporate the private image gains from contributing, even when in the aggregate these are offset by the image losses of non-contributors.[16]

[14] A related point is made by Chen and Schonger (2016b) for other forms of preferences involving moral "duties" in addition to material payoffs.

[15] It is also interesting (and somewhat reassuring) to compare our results to those of Falk and Szech (2013), in which subjects could accept money or forfeit it to save the life of one (surplus) mouse. There were four uniform price-list treatments, offering 20 choices each (in steps of 2.50€ or 5€), with two price ranges and two framings of the decision context. The fractions of "observationally deontological" participants were 18.8% and 27.1% (with $n = 192$) for a price support of [2.50€, 5€], and still 7.4% and 11.1% (with $n = 248$) for [5€, 100€]. The price-list data were used to calibrate parameters in that paper's main experiment, and are available upon request.

[16] In the model, the two exactly offset each other in the aggregate, making reputation a purely positional good. This is a (convenient) benchmark case, as empirically the game could have negative, zero, or positive sum, depending on whether the value of reputation is a concave, linear or convex functional. Butera et al. (2019) find evidence for negative sum, which reinforces the point made above.

Several directions for further research seem potentially interesting. First, both in the model and in the experiment, actors and observers (e.g., the "committee") had common knowledge of all aspects of the tradeoffs involved. In practice, the personal and social consequences of a selfish or prosocial act are often uncertain, and parties will exchange arguments over how costly, important, or unimportant, it is. Bénabou et al. (2018) therefore expand the basic prosociality-and-signaling framework to formalize the search for, and the social dissemination of moral *arguments:* excuses for pursuing self interest, or reasons why one (typically, someone else) should not do so. Bartling et al. (2020) study such strategic disclosures experimentally, while Foerster and van der Weele (2018a,b) independently explore similar questions in contexts where communication takes the form of cheap talk.

Second, we focussed here on people's (un)willingness to engage in moral tradeoffs that involve weighing their private costs or benefits (c and $\mu\hat{v}$) from some course of action against its social consequences (e). Another very important type concerns "trolley-like" choices between alternative public goods (e vs. e'), in which departing from some default action causes harm ($-e$) to some people (or even animals) but can be a means to achieve a greater good, such as benefiting or saving many more ($e' > e$). This is the focus of ongoing work.

References

Ambuehl, Sandro (2017). "An offer you can't refuse? Incentives change how we inform ourselves and what we believe". CESifo Working Paper No. 6296.

Andreoni, James (1989). "Giving with impure altruism: Applications to charity and Ricardian equivalence". *Journal of Political Economy*, 97.6, pp. 1447-1458.

— (1990). "Impure altruism and donations to public goods: A theory of warm-glow". *The Economic Journal*, 100.401, pp. 464-477.

Ariely, Dan, Anat Bracha, and Stephan Meier (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially". *American Economic Review*, 99.1, pp. 544-555.

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

Bartling, Björn, Valero, Vanessa, Weber, R. and Lan Yao (2020) "Public discourse and socially responsible market behavior", August mimeo.

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, pp. 17-34.

Bénabou, Roland, Armin Falk, and Jean Tirole (2018). "Narratives, imperatives, and moral persuasion", NBER Working Paper 24798.

Bénabou, Roland and Jean Tirole (2006). "Incentives and prosocial behavior". *American Economic Review*, 96.5, pp. 1652-1678.

— (2011). "Identity, morals, and taboos: Beliefs as assets". *Quarterly Journal of Economics*, 126.2, pp. 805-855.

Bentham, Jeremy (1789). *An introduction to the principles of morals.* London: Athlone.

Bos, Olivier and Martin Pollrich (2020) "Optimal auctions with signaling bidders", mimeo.

Bos, Olivier and Tom Truyts (2019) "Entry in first-price auctions with signaling", mimeo.

Brandts, Jordi and Gary Charness (2011). "The strategy versus the direct-response method: A first survey of experimental comparisons". *Experimental Economics*, 14.3, pp. 375-398.

Brekke, Kjell A., Snorre Kverndokk, and Karine Nyborg (2003). "An economic model of moral motivation". *Journal of Public Economics*, 87.9-10, pp. 1967-1983.

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, pp. 415-437.

Bursztyn, Leonardo, Gregory Egorov, and Stefano Fiorin (2018). "From extreme to mainstream: The erosion of social norms" *American Economic Review,* forthcoming.

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky (2019). "The deadweight loss of social recognition," U.C. Berkeley mimeo, March.

Charness, Gary, Uri Gneezy, and Brianna Halladay (2016). "Experimental methods: Pay one or pay all". *Journal of Economic Behavior and Organization*, 131, pp. 141-150.

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

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

Cox, James C., Vjollca Sadiraj, and Ulrich Schmidt (2015) Paradoxes and mechanisms for choice under risk," Experimental Economics 18(2), pp.. 215-250.

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.

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

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, pp. 88-97.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier (2012). "Testing for altruism and social pressure in charitable giving". *Quarterly Journal of Economics*, 127.1, pp. 1-56.

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao (2019). "Estimating social preferences and gift exchange with a piece-rate design," U.C. Berkeley mimeo, June.

Elias, Julio, Mario Macis, and Nicola Lacetera (2016). "Efficiency-morality trade-offs in repugnant transactions: A choice experiment". IZA Discussion Paper No. 10187.

Ellingsen, Tore and Magnus Johannesson (2008). "Pride and prejudice: The human side of incentive theory". *American Economic Review*, 98.3, pp. 990-1008.

Exley, Christine L. (2016). "Excusing selfishness in charitable giving: The role of risk". *Review of Economic Studies*, 83.2, pp. 587-628.

Falk, Armin (2019) "Facing yourself - A note on self-image". Bonn University mimeo

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, pp. 1645-169.

Falk, Armin and Thomas Graeber (2020). "Negative long-run effects of prosocial spending on happiness". *Proceedings of the National Academy of Sciences*, 201914324.

Falk, Armin, Thomas Neuber, and Nora Szech (2020). "Diffusion of being pivotal and immoral outcomes". *Review of Economic Studies*, forthcoming.

Falk, Armin and Nora Szech (2013). "Morals and markets". *Science*, 340.6133, pp. 707-711.

Feddersen, Timothy, Sean Gailmard, and Alvaro Sandroni (2009). "Moral bias in large elections: Theory and experimental evidence". *American Political Science Review*, 103.2, pp. 175-192.

Foerster, Manuel and Joel van der Weele (2018a) "Denial and alarmism in collective action problems," Tinbergen W.P. University of Amsterdam, February.

Foerster, Manuel and J. van der Weele (2018b) "Persuasion, justification, and the communication of social impact," Tinbergen Institute W.P. 2018-067/I, August.

Gino, Francesca, Michael I. Norton, and Roberto A. Weber (2016). "Motivated bayesians: Feeling moral while acting egoistically". *Journal of Economic Perspectives*, 30.3, pp. 189-212.

Giovannoni, Francesco and Miltiadis Makris (2014), "Reputational bidding", *International Economic Review,* 55: 693-710.

Gneezy, Uri, Elizabeth A. Keenan, and Ayelet Gneezy (2014). "Avoiding overhead aversion in charity". *Science* 346.6209, pp. 632-635.

Goeree, Jacob K. (2003) "Bidding for the Future: Signaling in Auctions with an Aftermarket", *Journal of Economic Theory,* 108: 345-364.

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, pp. 255-276.

Greiner, Ben (2004). "An online recruitment system for economic experiments for economic experiments". *Forschung und wissenschaftliches Rechnen*, ed. by Kurt Kremer and Volker Macho. 63rd ed., pp. 79-93.

Grossman, Zachary and Joel J. van der Weele (2017). "Self-image and willful ignorance in social decisions". *Journal of the European Economic Association*, 15.1, pp. 173-217.

Hörner, Johannes and Nicolas Sahuguet (2007) "Costly Signaling in Auctions", *Review of Economic Studies,* 74: 173-206.

Kagel, John H. and Alvin E. Roth (1995). *The handbook of experimental economics.* Princeton: Princeton University Press.

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

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, pp. 81-86.

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.

Mazar, Nina, On Amir, and Dan Ariely (2008). "The dishonesty of honest people: A theory of self-concept maintenance". *Journal of Marketing Research*, 45.6, pp. 633-644.

Mill, John Stuart (1863). *Utilitarianism.* Ed. by Son & Bourn Parker. 1 ed. London: West Strand.

Sandel, Michael J. (2012). *What money can't buy: The moral limits of markets.* Farrar, Straus and Giroux.

Small, Deborah A., and George Loewenstein (2003). "Helping a victim or helping the victim: Altruism and identifiability," *Journal of Risk and Uncertainty,* 26(1): 5-16.

Small, Deborah A., George Loewenstein, and Paul Slovic (2007). “Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims,” *Organizational Behavior and Human Decision Processes*, 102, 143-153.

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.

Tiemersma, Edine W., Marieke J. van der Werf, Martien W. Borgdor, 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.

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:

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

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. ■

Single-crossing condition for the comparison of DE and MPL elicitation. Proposition 3 showed that, for any cost distribution $G(c)$, $\bar{a}^{DE}(\mu)$ and $\bar{a}^{MPL}(\mu)$ must cross at least one as μ rises from small positive values to sufficiently large ones. Focusing here on the uniform distribution almost universally used in experimental work, we provide, for the sake of completeness, a set of sufficient conditions for this intersection to be unique.

1. Direct elicitation. Equations (2)-(4) can be rewritten as:

$$c_H^{DE} = v_H e + \mu(1-\rho)(v_H - v_L),$$
$$\bar{c}_L^{DE} = v_L e + \mu(v_H - v_L),$$
$$\underline{c}_L^{DE} = v_L e + \mu\rho(v_H - v_L),$$

while the low types' mixed strategy for $c \in [\underline{c}_L^{DE}, \min\{\bar{c}_L^{DE}, c_H^{DE}\})$ is

$$a_L(c) = \frac{\rho}{1-\rho}\left[\frac{\mu(v_H - v_L)}{c - v_L e} - 1\right].$$

Substituting into (6) for a uniform distribution and focusing on the case where $\underline{c}_L^{DE} \leq c_H^{DE} \leq c_{\max}$ yields

$$\bar{a}^{DE} = \rho\frac{c_H^{DE}}{c_{\max}} + (1-\rho)\frac{\underline{c}_L^{DE}}{c_{\max}} + \int_{\underline{c}_L^{DE}}^{\bar{c}_L^{DE}} (1-\rho)a_L(\tilde{c})\frac{d\tilde{c}}{c_{\max}} \quad \mu < e/\rho,$$
$$\bar{a}^{DE} = \rho\frac{c_H^{DE}}{c_{\max}} + (1-\rho)\frac{\underline{c}_L^{DE}}{c_{\max}} + \int_{\underline{c}_L^{DE}}^{c_H^{DE}} (1-\rho)a_L(\tilde{c})\frac{d(\tilde{c})}{c_{\max}} \quad \text{for } \mu \geqslant e/\rho.$$

Hence:

$$\bar{a}^{DE}(\mu) = \begin{cases} \frac{1}{c_{\max}} \left[\bar{v}e + \mu\rho(v_H - v_L)(1-\rho-\log\rho)\right], & \text{if } \mu < e/\rho \\ \frac{1}{c_{\max}} \left[v_L e + \mu\rho(v_H - v_L)\left(1+\log\left(\frac{e}{\mu}+1-\rho\right)-\log\rho\right)\right], & \text{if } \mu \geqslant e/\rho \end{cases} . \| \tag{A.2}$$

2. Multiple Price List. With the uniform distribution $L(c) = (c - v_L e)^2/(2c_{\max})$, so the three cutoffs for μ defined in Proposition 2 are given by:

$$\underline{\mu} = \frac{(v_H - v_L)e^2}{2c_{\max}}, \quad \mu^* = \frac{1}{2c_{\max}} \frac{(c_{\max} - v_L e)^2}{v_H - v_L}, \quad \bar{\mu} = \frac{\mu^*}{\rho}, \tag{A.3}$$

and the two types' willingnesses to accept equal:

$$c_H^{MPL}(\mu) = \begin{cases} v_H e, & \text{if } \mu < \underline{\mu} \\ v_L e + \sqrt{2c_{\max}\mu(v_H - v_L)}, & \text{if } \mu \in [\underline{\mu}, \mu^*) \\ c_{\max}, & \text{if } \mu \geqslant \mu^* \end{cases},$$

$$c_L^{MPL}(\mu) = \begin{cases} v_L e, & \text{if } \mu < \mu^* \\ \begin{cases} v_L e, & \text{w.p. } a_L(\mu) \\ c_{\max}, & \text{w.p. } 1 - a_L(\mu) \end{cases}, & \text{if } \mu \in [\mu^*, \bar{\mu}) \\ c_{\max}, & \text{if } \mu \geqslant \bar{\mu} \end{cases} \tag{A.4}$$

Substituting into (7), we have

(a) If $\mu < \underline{\mu}$, then

$$\bar{a}^{MPL} = \rho\frac{v_H e}{c_{\max}} + (1-\rho)\frac{v_L e}{c_{\max}} = \frac{\bar{v}e}{c_{\max}}.$$

(b) If $\mu \in [\underline{\mu}, \mu^*)$, then

$$\bar{a}^{MPL} = \rho\frac{(v_L e + \sqrt{2c_{\max}\mu(v_H - v_L)})}{c_{\max}} + (1-\rho)\frac{v_L e}{c_{\max}} = \frac{v_L e}{c_{\max}} + \rho\sqrt{\frac{2\mu(v_H - v_L)}{c_{\max}}}$$

(c) If $\mu \in [\mu^*, \bar{\mu})$, then

$$\bar{a}^{MPL} = \rho + (1-\rho)\left[a_L(\mu) + (1 - a_L(\mu))\frac{v_L e}{c_{\max}}\right] = \frac{\mu}{\bar{\mu}} + \left(1 - \frac{\mu}{\bar{\mu}}\right)\frac{v_L e}{c_{\max}},$$

since $a_L(\mu) = (\mu - \rho\bar{\mu}) / [(1-\rho)\bar{\mu}]$

For $\mu > \bar{\mu}$, finally, we saw that $\bar{a}^{MPL} = 1$. Summarizing, we have

$$\bar{a}^{MPL}(\mu) = \begin{cases} \frac{\bar{v}e}{c_{\max}}, & \text{if } \mu < \underline{\mu} \\ \frac{v_L e}{c_{\max}} + \rho\sqrt{\frac{2\mu(v_H - v_L)}{c_{\max}}}, & \text{if } \mu \in [\underline{\mu}, \mu^*) \\ \frac{\mu}{\bar{\mu}} + \left(1 - \frac{\mu}{\bar{\mu}}\right)\frac{v_L e}{c_{\max}}, & \text{if } \mu \in [\mu^*, \bar{\mu}) \\ 1, & \text{if } \mu \geqslant \bar{\mu} \end{cases} \quad \|$$

3. Auxiliary assumptions. We will focus for the moment on values $\mu \leq 1/\rho$ (the maximum value for which Assumption 1 is feasible for some e). Recall that, in computing $\bar{a}^{DE}$ we assumed that $c_{\max} \geqslant c_H^{DE}(\mu)$. Since c_H^{DE} is increasing, we need only impose this at $\mu = 1/\rho$, which means:

$$c_{\max} \geqslant v_H e + \frac{1-\rho}{\rho}(v_H - v_L). \tag{A.5}$$

Second, $\bar{a}^{MPL}(\mu) = 1 > \bar{a}^{DE}(\mu)$ for $\mu \geqslant \bar{\mu}$, so we need only look for intersections at $\mu < \bar{\mu}$. The above range restriction for μ then requires that $\bar{\mu} \leq 1/\rho$ (equivalently, $\mu^* \leq 1$), that is:

$$(c_{\max} - v_L e)^2 \leqslant 2c_{\max}(v_H - v_L). \tag{A.6}$$

Finally, we also imposed that $c_H^{DE} \geq \underline{c}_L^{DE}$, which means that

$$e + \mu(1-\rho) > \mu\rho. \tag{A.7}$$

Thus, it must be that either $\rho \leqslant 1/2$, or else $\rho > 1/2$ and $e \geqslant 2 - 1/\rho$.

With these three assumptions, it is always the case that $\underline{\mu} < e/\rho$, but we still have three possible cases for the remaining cutoffs: $\underline{\mu} < e/\rho < \mu^* < \bar{\mu}$, $\underline{\mu} < \mu^* < e/\rho < \bar{\mu}$, and $\underline{\mu} < \mu^* < \bar{\mu} < e/\rho$. $\|$

4. Single-Crossing Condition. We will show that, together with (A.5)–(A.7) above, the following condition ensures that $\bar{a}^{MPL}(\mu)$ and $\bar{a}^{DE}(\mu)$ cross only once:

$$e + \mu^*(1 - \rho - \log\rho) < 2\mu^* \frac{c_{\max}}{c_{\max} - v_L e} = \frac{c_{\max} - v_L e}{\rho} \tag{A.8}$$

It will be useful to define $V_L \equiv v_L e/c_{\max}$ and $V_\Delta = (v_H - v_L)/c_{\max}$, and then from these:

$$A^{DE}(\mu) \equiv \frac{\bar{a}^{DE}(\mu) - V_L}{\rho V_\Delta} = \begin{cases} e + \mu(1 - \rho - \log\rho), & \text{if } \mu < e/\rho \\ \mu\left(1 + \log\left(\frac{e}{\mu} + 1 - \rho\right) - \log\rho\right), & \text{if } \mu \geqslant e/\rho \end{cases}, \tag{A.9}$$

$$A^{MPL}(\mu) \equiv \frac{\bar{a}^{MPL}(\mu) - V_L}{\rho V_\Delta}. = \begin{cases} e, & \text{if } \mu < \underline{\mu} \\ \sqrt{\frac{2\mu}{V_\Delta}}, & \text{if } \mu \in [\underline{\mu}, \mu^*) \\ \frac{2\mu}{1 - V_L}, & \text{if } \mu \in [\mu^*, \bar{\mu}] \end{cases}. \tag{A.10}$$

The MPL cutoffs are thus given by

$$\underline{\mu} = \frac{V_\Delta e^2}{2}, \quad \mu^* = \frac{1}{2V_\Delta}(1-V_L)^2, \quad \bar{\mu} = \frac{1}{2\rho V_\Delta}(1-V_L)^2, \tag{A.11}$$

and Assumptions (A.5)–(A.7) can be rewritten as:

$$V_L + V_\Delta \left(e + \frac{1-\rho}{\rho} \right) \leqslant 1, \tag{A.12}$$

$$(1-V_L)^2 \leqslant 2V_\Delta, \tag{A.13}$$

$$\rho \leqslant 1/2, \text{ or } \rho > 1/2 \text{ and } e \geqslant 2 - 1/\rho. \tag{A.14}$$

As to the single-crossing condition, it takes the form.

$$e + \mu^*(1-\rho-\log\rho) < 2\mu^* \frac{1}{1-V_L}. \tag{A.15}$$

Let us first show that when it holds, then

$$A^{DE}(\mu^*) \leqslant e + \mu^*(1-\rho-\log\rho) < A^{MPL}(\mu^*).$$

If $\mu^* \leqslant e/\rho$, this follows from the definition of A^{DE}. If $\mu^* > e/\rho$, then note that

$$A'^{DE}(\mu) = \begin{cases} 1-\rho-\log\rho, & \text{if } \mu < e/\rho \\ 1 + \log\left(\frac{e}{\mu} + 1 - \rho\right) - \log\rho - \frac{e}{e+\mu(1-\rho)}, & \text{if } \mu \geqslant e/\rho, \end{cases},$$

and that these left- and right-derivatives coincide at e/ρ. Moreover, for $\mu > e/\rho$ we have $A''^{DE} = -e^2/\left[\mu(e+\mu(1-\rho))^2\right] < 0$, and therefore $A^{DE}(\mu) < e + \mu(1-\rho-\log\rho)$.

Next, note that (A.15) implies that $e+\mu(1-\rho-\log\rho)$ and $A^{MPL}(\mu)$ cross exactly once for $\mu \in [0, \mu^*]$, since the first function is linear and the second concave, and we know that for $\mu < \underline{\mu}$, $A^{DE} > A^{MPL}$. By the previous bounding argument, it follows that $A^{DE}(\mu)$ and $A^{MPL}(\mu)$ also cross exactly once in this region. Finally, (A.15) also implies that

$$\frac{1-V_L}{2}(1-\rho-\log\rho) < 1 - \frac{V_\Delta e}{1-V_L} < 1,$$

so $A'^{MPL}(\mu) > 1-\rho-\log\rho \geqslant A'^{DE}(\mu)$, for $\mu > \mu^*$. Therefore, $A^{MPL}(\mu) > A^{DE}(\mu)$ for all $\mu > \mu^*$. Thus, condition (A.15) indeed guarantees that A^{DE} and A^{MPL} (and so $\bar{a}^{DE}$ and $\bar{a}^{MPL}$) cross exactly once, and we can find μ^* such that $\bar{a}^{MPL} > \bar{a}^{DE}$ iff $\mu > \mu^*$.‖

5. Compatibility of the four conditions (for $\rho \leqslant 1/2$). We now verify that the intersection of (A.12)–(A.14) and the single-crossing condition (A.15) define a nonempty region region of parameters.

First, we can always find $0 < \rho \leqslant 1/2$ (ensuring (A.14)), $e \in [0,1]$ and $V_L \in [0,1]$ such that

$$e - \rho + \frac{1}{\rho} < \frac{2}{1-V_L}.$$

Then, take $x \in (e-\rho+1/\rho, 2/(1-V_L))$ and let $V_\Delta = (1-V_L)/x$. Condition (A.13) then holds, since

$$\frac{(1-V_L)^2}{2V_\Delta} = x\frac{1-V_L}{2} < 1,$$

and similarly for (A.12), since

$$e + \frac{1-\rho}{\rho} < e - \rho + \frac{1}{\rho} < x = \frac{1-V_L}{V_\Delta}.$$

The single-crossing condition (A.15), finally, requires that

$$e + \mu^*(1-\rho-\log\rho) = e + \frac{(1-V_L)^2}{2V_\Delta}(1-\rho-\log\rho) < e + 1 - \rho - \log\rho.$$

But $\log\rho > 1 - 1/\rho$, so

$$e + \mu^*(1-\rho-\log\rho) < e - \rho + \frac{1}{\rho} < x = \frac{1-V_L}{V_\Delta} = \frac{2\mu^*}{1-V_L}.$$

All conditions are then simultaneously satisfied.

6. Extending the results to $\mu > 1/\rho$. Note that in this case in may be that $c_L^{DE}(\mu) > \underline{c}_H^{DE}(\mu)$. We will show that (A.15) alone guarantees single crossing. First, we establish that, for all μ :

$$A^{DE}(\mu) \leqslant e + \mu(1-\rho-\log\rho). \tag{A.16}$$

Defining μ_H and μ_L, respectively, by $c_H^{DE}(\mu_H) \equiv c_{\max}$ and $\underline{c}_L^{DE}(\mu_L) \equiv c_{\max}$, there are two cases to consider.

(a) Case $\mu_H < \mu_L$. The function $A^{DE}(\mu)$ is given by:

$$A^{DE}(\mu) = \begin{cases} e + \mu(1-\rho-\log\rho), & \text{if } \mu \leqslant e/\rho \\ \mu\left(1 + \log\left(\frac{e}{\mu} + 1 - \rho\right) - \log\rho\right), & \text{if } e/\rho < \mu \leqslant \mu_H \\ \mu\left(1 + \log\left(\frac{1-V_L}{\rho V_\Delta}\right) - \log\mu\right), & \text{if } \mu_H < \mu \leqslant \mu_L \\ \frac{1-V_L}{\rho V_\Delta}, & \text{if } \mu > \mu_L \end{cases},$$

with first derivative

$$A'^{DE}(\mu) = \begin{cases} 1-\rho-\log\rho, & \text{if } \mu \leqslant e/\rho \\ 1 + \log\left(\frac{e}{\mu} + 1 - \rho\right) - \log\rho - \frac{e}{e+\mu(1-\rho)}, & \text{if } e/\rho < \mu \leqslant \mu_H \\ \log\left(\frac{1-V_L}{\rho V_\Delta}\right) - \log\mu, & \text{if } \mu_H < \mu \leqslant \mu_L \\ 0, & \text{if } \mu > \mu_L \end{cases}$$

Note that $A^{DE}(\mu)$ is continuous, but $A'^{DE}(\mu)$ is discontinuous at μ_H. In particular, as long as

$v_H e < c_{\max}$,

$$\lim_{\mu \nearrow \mu_H} A'^{DE}(\mu) > \lim_{\mu \searrow \mu_H} A'^{DE}(\mu).$$

Also, it is still the case that $A''^{DE}(\mu) \leqslant 0$. Combining everything, we can see that (A.16) indeed holds for all μ.

(b) Case $\mu_L < \mu_H$. Define $\bar{\mu}$ by $\underline{c}_L^{DE}(\bar{\mu}) \equiv \underline{c}_H^{DE}(\bar{\mu})$, which implies that $\bar{\mu} < \mu_L < \mu_H$. Note that this requires that $\rho > 1/2$.

The function $A^{DE}(\mu)$ is given by:

$$A^{DE}(\mu) = \begin{cases} e + \mu(1 - \rho - \log \rho), & \text{if } \mu \leqslant e/\rho \\ \mu \left(1 + \log\left(\frac{e}{\mu} + 1 - \rho\right) - \log \rho\right), & \text{if } e/\rho < \mu \leqslant \bar{\mu} \\ \frac{e+\mu(1-\rho)}{\rho}, & \text{if } \bar{\mu} < \mu \leqslant \mu_H \\ \frac{1-V_L}{\rho V_\Delta}, & \text{if } \mu > \mu_H \end{cases},$$

with first derivative:

$$A'^{DE}(\mu) = \begin{cases} 1 - \rho - \log \rho, & \text{if } \mu \leqslant e/\rho \\ 1 + \log\left(\frac{e}{\mu} + 1 - \rho\right) - \log \rho - \frac{e}{e+\mu(1-\rho)}, & \text{if } e/\rho < \mu \leqslant \bar{\mu} \\ \frac{1-\rho}{\rho}, & \text{if } \bar{\mu} < \mu \leqslant \mu_H \\ 0, & \text{if } \mu > \mu_H \end{cases}$$

Thus $A''^{DE}(\mu) \leqslant 0$ and $A'^{DE}(\mu)$ is continuous at $\bar{\mu}$, so again (A.16) holds for all μ.

Finally, recall that we showed (A.15) to imply that $e + \mu^*(1 - \rho - \log \rho) < A^{MPL}(\mu^*)$, which in turn ensures that $A'^{MPL}(\mu) > 1 - \rho - \log \rho$ for $\mu > \mu^*$. Therefore

$$A^{DE}(\mu) \leqslant e + \mu(1 - \rho - \log \rho) < A^{MPL}(\mu), \text{ for all } \mu > \mu^*,$$

ensuring again that A^{DE} and A^{MPL} cross at a single point, which lies in $[0, \mu^*]$.■

Appendix B: Additionnal Tables

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
R^2	0.030	0.078

The table shows OLS regression coefficients. Robust standard errors in parentheses. Controls include age, gender, income, religiousness, educational level, and high school grade. Significance levels: *p<0.1, **p<0.05 and ***p<0.01.

Appendix C: Instructions

C.1 Announcement by the Experimenter

The following text was read aloud by the experimenter after all subjects were placed in their cubicles. The content depended on the image treatment.

C.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.

C.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.

C.2 Introduction

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

C.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".

C.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.

C.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.

Figure 3: Typical appearance of a tuberculosis patient.

Operation ASHA

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 4: 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 illustrated in the following graphic: [Figure 5 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

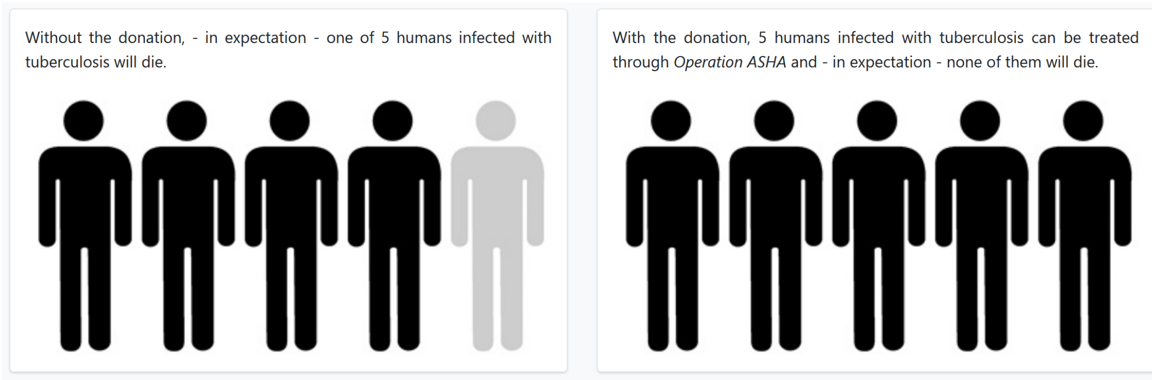

Figure 5: Relationship between the donation and the saving of a life

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.

C.3 Treatment *DE Low Image*

C.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.

C.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.

C.4 Treatment *DE High Image*

C.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.

C.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.

C.5 Treatment *MPL Low Image*

C.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.

C.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.

C.6 Treatment *MPL High Image*

C.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.

C.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.

C.7 Decision Screens

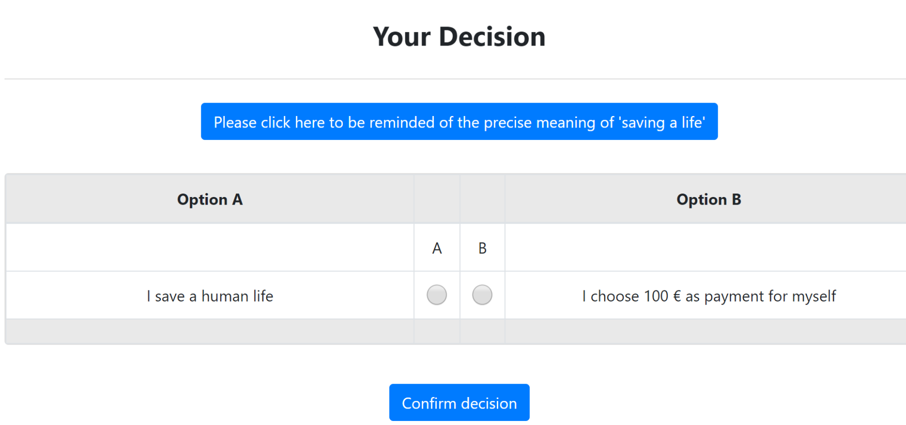

Figure 6: Decision Screen DE

Your Decisions

Please click here to be reminded of the precise meaning of 'saving a life'

Option A				Option B
	A		B	
I save a human life	○	1	○	I choose 0 € as payment for myself
I save a human life	○	2	○	I choose 10 € as payment for myself
I save a human life	○	3	○	I choose 20 € as payment for myself
I save a human life	○	4	○	I choose 30 € as payment for myself
I save a human life	○	5	○	I choose 40 € as payment for myself
I save a human life	○	6	○	I choose 50 € as payment for myself
I save a human life	○	7	○	I choose 60 € as payment for myself
I save a human life	○	8	○	I choose 70 € as payment for myself
I save a human life	○	9	○	I choose 80 € as payment for myself
I save a human life	○	10	○	I choose 90 € as payment for myself
I save a human life	○	11	○	I choose 100 € as payment for myself
I save a human life	○	12	○	I choose 110 € as payment for myself
I save a human life	○	13	○	I choose 120 € as payment for myself
I save a human life	○	14	○	I choose 130 € as payment for myself
I save a human life	○	15	○	I choose 140 € as payment for myself
I save a human life	○	16	○	I choose 150 € as payment for myself
I save a human life	○	17	○	I choose 160 € as payment for myself
I save a human life	○	18	○	I choose 170 € as payment for myself
I save a human life	○	19	○	I choose 180 € as payment for myself
I save a human life	○	20	○	I choose 190 € as payment for myself
I save a human life	○	21	○	I choose 200 € as payment for myself

Confirm decisions

Figure 7: Decision Screen MPL