Cheap Signaling of Altruism*

Moritz Janas
Center for Behavioral Institutional Design
New York University Abu Dhabi
PO Box 129188
Abu Dhabi
United Arab Emirates
moritz.janas@nyu.edu

Michelle Jordan
Department of Economics
University of Konstanz
PO Box 150
78457 Konstanz
Germany
michelle.jordan@uni-konstanz.de

March 1, 2024

Abstract

Decisions on whether to engage in a pro-social act can separate altruistic from non-altruistic individuals. We explore the role of the probability that the pro-social action has to be carried out after publicly deciding in favor of it. In such a signaling environment, a lower probability that the act has to be carried out *cheapens* the signal. We use a model to predict how this cheapness influences decision-making behavior and the updating of beliefs about the decision-maker's level of altruism. In a laboratory experiment, we test the model's predictions by varying the probability that the pro-social decision has to be carried out and the strength of image concerns. If the image concern is non-monetary, the experimental data reveals that, in line with the model's predictions, the share of pro-social decisions increases in the case of cheaper signals. The prediction that the effect of the cheapness increases with image concerns cannot be confirmed by the data. Belief-updating is a crucial element of modeling pro-social decision-making in signaling environments, as it provides the trade-off between costs and image benefits. However, the experimental data does not show significant belief-updating differences depending on the cheapness of the signal after observing a pro-social decision.

Keywords: signaling games, altruism, philanthropy, pivotality, laboratory experiment.

JEL: C91, C92, D64, D82, D83.

*We would like to thank Urs Fischbacher, Susanne Goldlücke, the TWI research group, and the participants at several workshops and conferences for their valuable comments. All remaining errors are our own. The authors gratefully acknowledge funding by the Graduate School of Decision Sciences at the University of Konstanz and Moritz Janas gratefully acknowledges financial support from Tamkeen under NYU Abu Dhabi Research Institute Award CG005. Data, experimental materials, and code to replicate the analysis can be accessed from the OSF repository at the link osf.io/ajyct. Corresponding author: Moritz Janas. The authors declare no conflicts of interest.

1 Introduction

Observing someone helping a friend, donating money, or volunteering for a benevolent cause can tell us something about the kindness of this person. The level of kindness cannot be directly observed, but such behavior can serve as an informative signal. If people who are kinder than others derive a higher intrinsic benefit from pro-social actions, some pro-social actions will only be done by kind individuals. Observing a person carrying out such a costly action reveals that this person is kind.
In reality, however, many pro-social actions may turn out to be no longer necessary, and thus committing to them can become costless after the decision to take the action. What one can learn from observing the individual in a situation where the signaling of kindness becomes *cheaper* is not always straightforward. One example is conditional favors. Suppose, e.g., an acquaintance asks you whether he can stay a few nights at your place if he is accepted at a specific conference. You both know, however, that there is only a particular likelihood of him being accepted. So, even if you assure him of your hospitality, you might not have to host your acquaintance, and thus you might not have any costs. Another example is volunteering to referee a paper. For many people, the costs usually outweigh the benefits; therefore, only people who are kind enough to spend time reading and commenting on the paper accept a referee invitation. Every now and then, however, after agreeing to review the article, the editor notifies a referee that the report is no longer needed. In such a situation, the *kind* decision to accept the referee invitation, which is observable to the editor, has a small but positive probability of not incurring any costs. An example with more severe consequences is stem cell donation. Being tested and added to the database can be seen as kind behavior. However, the probability of helping someone (and incurring the costs) depends on the exogenous probability that a match is found. In this paper, we study such situations. We explore a signaling game that can be used to transmit a signal about the unobservable level of altruism to others by engaging in a costly act. We test how such conditionality changes decision-making by varying the publicly known probability that the computer will carry out a decision in favor of the pro-social option. Further, as the belief-updating after observing an action is a crucial part of the equilibrium in such models, we also explore how such a change in the game changes how others interpret pro-social behavior in each situation.

Although people frequently encounter situations in which signaling altruism is cheap, only a few papers have studied such decision contexts (e.g., Grossman, 2015; Carter and Guerette, 1992; Fischer, 1996; Tyran, 2004). They all predict that more people should decide pro-socially when signaling altruism is cheaper. However, the experimental findings are mixed: some fail to confirm this prediction. One potential reason may be that the prediction only works for certain types of image concerns. Carter and Guerette (1992) use a design with complete anonymity and find a weak effect of reduced pivotality for some parameter values, and no effect of reduced pivotality for other values. Fischer (1996) find that more people vote in favor of the pro-social option if anonymity is reduced and, in anonymous voting, if the decision's pivotality is reduced. Tyran (2004) uses a double-blind procedure and finds the opposite effect: when reducing the pivotality of the decision, more people choose the pro-social option. Experiment 2 in Grossman (2015) finds no effect of reduced pivotality for an anonymous decision, and finds a positive effect of reduced pivotality on pro-social decisions in an environment with the experimenter as the audience. To see how far image concerns can explain these mixed findings, we also systematically vary the type of image incentives in our experiment. Moreover, we elicit how others perceive decisions when signaling altruism is cheap compared to when it is not. So far, perceptions have received no direct attention. Yet, all of the above-mentioned papers assume (partly indirectly) that people have image concerns, i.e., that they care about how their decisions are perceived by others (or themselves). Thus, it is only natural to ask how others perceive such decisions. Our paper aims to contribute to belief formation in cheap signaling situations.

Overall, we address the following questions: First, do more people decide pro-socially if the (costly) pro-social act is less likely to be enacted, i.e., if signaling altruism is cheaper? Second, how does the answer to this question depend on the strength and type of image concerns? Third, do observers realize that donation decisions are less diagnostic and selfish decisions are more diagnostic if signaling altruism is cheaper?

We first use a model based on Bénabou and Tirole (2006) and Grossman (2015) to derive comparative statics predictions for how behavior and perceptions should change when the signaling becomes cheaper, and for how the change in decisions and perceptions should depend on the image concern incentives. We then test these predictions in a laboratory experiment. In the experiment, subjects can decide between keeping money or donating to a charity. This choice is made observable to another subject. However, with a positive probability, donation turns out to be not executed ex-post, i.e., after the decision-maker has made his choice. We manipulate the type of image concerns between subjects. Choices are made observable either in an anonymous manner, with a picture of the decision-maker, or with a picture and a monetary bonus for a pro-social appearance. To elicit the perceptions of a decision-maker after an action, a decision-maker in our experiment also decides privately whether to donate money to charity. An observer, who observes the public decision, has to enter their beliefs about the likelihood that the decision-maker will donate in the private situation. Knowing the real decision in the private situation allows us to incentivize the observer for the accuracy of their belief. We can draw inferences on the first and third question by comparing decisions and perceptions in settings with different signaling costs. Comparing decisions and perceptions across different image concern incentives gives us insights into how to answer the second question.

Our experimental findings can be summarized as follows. As postulated by the model, we find that more individuals decide pro-socially if signaling altruism is cheaper. However, this is only true if the image concern incentive is low and non-monetary. In the treatment where a pro-social appearance earns decision-makers a monetary bonus, the fraction of pro-social decisions does not increase in response to cheapened signaling costs. The model predicts that cheapened signaling costs should matter more when image concerns are stronger. The experimental data, however, does not allow us to reject the hypothesis that the cheapened signaling costs matter to the same degree in this case. Moreover, the model predicts that observers account for the cheapness of signaling by discounting the signals. For cheaper signals, donation decisions should lead to lower altruism estimates. Similarly, decisions against donation should also lead to lower altruism estimates of the observer. Our experimental data, however, finds a systematic bias: observers' perceptions responded relatively little to cheapened signaling costs after a decision in favor of donation, and even to a higher estimated level of altruism after a non-donation decision in the picture treatment. A power calculation suggests that the failure to reject the null hypothesis of no difference in updating after a donation decision for different levels of cheapness is unlikely to be driven by a small sample size. In theory, when deciding whether to donate, the decision-maker should trade off the costs and image gains of both options. If, however, the image gains do not follow theory, as observers fail to update their belief negatively in the case of a cheaper no-donation decision, the decision-maker's trade-off can become far more complicated than modeled by a Perfect Bayesian Equilibrium.

Closely related are papers that study pro-social (or moral or fair) decisions in a context in which not only image concerns but also pivotality plays a role (e.g. Carter and Guerette, 1992; Tyran, 2004; Grossman, 2015; Exley, 2016; Falk et al., 2020; Andreoni and Bernheim, 2009; Bénabou et al., 2022). However, only in Carter and Guerette (1992), Tyran (2004), Grossman (2015), and Bénabou et al. (2022) does the uncertainty that a costly pro-social decision will be implemented cheapen the signaling of altruism. The uncertainty in Exley (2016), Falk et al. (2020), and Andreoni and Bernheim (2009) creates an excuse to decide selfishly, making a pro-social decision less, and not more, attractive. Bénabou et al. (2022) compare the extent of pro-social behavior under two elicitation methods that differ in their expected pivotality. By varying the level of visibility of the subjects' decisions and the elicitation method, they find that, with a low level of visibility, the high pivotality elicitation mechanism leads to more pro-social outcomes than the low pivotality elicitation mechanism. With a high level of visibility, however, the low pivotality elicitation mechanism leads to more pro-social outcomes than the high pivotality elicitation.

The experimental papers most closely related to ours are Grossman (2015) and Fischer (1996). In Grossman (2015), subjects play a dictator game and can either be kind or selfish. Depending on the treatment, the decision is either likely or unlikely to be overruled by the computer. By varying whether the recipient can only observe the choice or also the outcome, Grossman tests whether subjects care more about their self-image or their social image. In an additional experiment in Grossman (2015), the receiver is a charity and the observer is the experimenter. The results in Grossman (2015) are a bit noisy, but by conducting a reasonable sub-sample analysis without selfish subjects, the results speak in favor of social- instead of self-signaling. We use the theoretical model by also assuming that the subjects' utility function depends on an expected monetary gain as well as an image gain determined by the beliefs of others. Our paper extends Grossman (2015) by i) manipulating the strength of the image concern towards other subjects instead of towards the experimenter, ii) a prediction and test of the interaction of reduced pivotality and image concerns, and iii) an incentive-compatible elicitation of observers' beliefs. Eliciting the observers' beliefs enables us to test whether observers update their beliefs in a Bayesian way, a crucial element of the equilibrium prediction.

Fischer (1996) conducts a classroom experiment in an expressive voting framework. By manipulating the pivotality and the visibility of a pro-social choice, he finds that some subjects only vote pro-socially if the vote is less likely to be pivotal and that subjects also react to the anonymity of the vote. Fischer (1996), however, explores neither whether the effect of reduced pivotality depends on the anonymity of the vote, nor how others perceive the votes.

Somewhat, but less closely, related are papers that use a signaling model to explain pro-social behavior but do not directly focus on pivotality (e.g. Tonin and Vlassopoulos, 2013; Ellingsen and Johannesson, 2011; Bénabou et al., 2022; Klinowski, 2021; Meyer and Tripodi, 2021). Papers in which signaling costs are cheapened by reduced pivotality (Carter and Guerette, 1992; Fischer, 1996; Tyran, 2004; Grossman, 2015; Ginzburg et al., 2022) focus on how the inclination to decide (vote) pro-socially changes when the pivotality of a pro-social decision (vote) is decreased. Evidence regarding the answer to this question is mixed. Carter and Guerette (1992) find weakly significant evidence for more donation votes in cases of less pivotality but cannot replicate this trend for different parameters. Fischer (1996) finds only a low share (24.5%) of subjects reacting to changes in pivotality but finds a significant increase in pro-social choices for lower pivotality among these subjects. Tyran (2004) finds evidence for a bandwagon effect, i.e., subjects behaving as they expect others to act, which contradicts the idea of lower approval for a pro-social decision in cases where the pivotality within a committee increases. The results in Grossman (2015) are in line with social signaling models, and provide little evidence for self-signaling motives. For a detailed literature review on expressive voting, see Tyran and Wagner (2019). Despite the theoretical importance of perceptions, these previous experiments do not explore how perceptions are affected by a decline in pivotality. We fill this gap. Moreover, by systematically varying the strength of social image concerns, we explore the role image concerns play in the context of reduced pivotality and offer a potential explanation for why previous evidence is mixed. Overall, we are not aware of any experimental paper that explicitly tests the perception of people's pro-social decisions when a) the signaling costs change, and/or b) the strength of the image concerns differs.

2 Experiment and hypotheses

2.1 Experimental design

In the experiment, there are two different roles: donor (he) and observer (she). The treatments differ in terms of their parameter values but have a similar general structure. Subjects start in the donor role. In this role, subjects can choose between donating money to a charity or keeping it. For 16 rounds, subjects face two such decisions per round. The first decision per round is *public* and uncertain. Subjects are informed that this first decision per round will be shown to another participant later on and if they decide in favor of donation, an exogenous probability p overrules a donation decision and makes subjects keep the money in this round. The second decision in each round is *private* and certain. Other subjects do not get to see the decision, and the decision cannot be overruled with an exogenous probability. For each decision, the *public* and the *private*, subjects get to choose between keeping 100 points or donating $X \times 100$ points to a charity with $X \in \{0.5, 0.75, 1.5, 2.0, 2.5, 3.0\}$. The value of the efficiency factor X is randomly drawn for each subject in each donation situation.

Throughout the 16 rounds, there is no interaction and no feedback is provided.
After these 16 rounds, subjects play the observer role. As observer, they state their beliefs about the behavior of other participants in the first 16 rounds. To form this belief, subjects observe the decision situation of one of the rounds of a randomly drawn other participant. This situation varies for different values of p and two values of X (the efficiency factor of the donation in the *public* and the *private* decision). After observing the decision situation, subjects state the probability that the decision-maker will choose to donate in the *private* setting given their decision in the *public* setting. This belief is elicited via strategy method: for both possible decisions in the *public* decision (whether to donate or to keep the money), observers enter a number between 0 and 100 via a slider. This number indicates the probability that one believes the observed subject will choose to donate in the *private* setting. This belief is incentivized via quadratic scoring rule for binary outcomes (see Nyarko and Schotter (2002), for a discussion, see Erkal et al. (2020); Folli and Wolff (2022); Burdea and Woon (2022)). If the observed subject chooses to donate in the *private* decision, the observer receives a payoff of $\Pi_{donated} = 100 - \frac{1}{100} \cdot (100 - belief)^2$ points. If the observed subject chooses not to donate, the observer's payoff is calculated as $\Pi_{notdonated} = 100 - \frac{1}{100} \cdot (belief)^2$. This payoff rule makes it incentive-compatible to state the true belief about the *private* decision. To make it easier to understand, instead of showing the formula on the observer's screen, the potential resulting payoffs of the stated beliefs are shown and updated in real time according to the number selected on the slider. An observer, therefore, sees, for example when stating a belief of 75%, that the resulting payoff will be 93.75 points in case the observed subject chooses to donate in the *private* decision, and 43.75 points in case the observed subject chooses to keep the money in the *private* decision.
In each of the 16 rounds where subjects act in the role of the observer, after entering the beliefs for both potential *public* decisions of the observed subject, the observer is informed about the decision-maker's *public* decision. After 16 rounds in the observer role, subjects answer a non-incentivized post-experimental questionnaire about their age, gender, field of study, attitude towards lying, income, and grades in high school.

After handing out the instructions but before the first round starts, subjects answer an understanding quiz about the incentive scheme and the game's main rules. At the end of the experiment, subjects choose in private whether their donation(s) should go to Amnesty International, Médecins sans frontières, or the German Cancer Aid. We have chosen charities that rank high in credibility and respectability surveys. We offer a list of charities to increase the chance that each subject will find at least one charity they perceive to be deserving. After choosing their preferred charity, subjects receive their payment in private, and we transfer the donated amount to the indicated charity after the session. The payment consists of a flat payment (5 Euro), two randomly selected estimates made as *observer*, and four decisions made as *decision-maker* (one randomly chosen *public* and *private* decision for rounds $1 - 8$ and rounds $9 - 16$ respectively). Points are converted with an exchange rate of 100 points = 1.5 Euro. By randomly selecting two rounds for the stated beliefs as an observer and not revealing which rounds are selected, we ensure that the observers' payoffs do not reveal the behavior of the observed subject in the *private* decision.

Treatments We have varied the strength of the image concerns between subjects. For this, we vary how *public* decisions are presented to observers after they enter their estimate. Overall, we have three different image concern treatments: *Anonymous*, *Picture*, and *Picture+Bonus*. In the *Anonymous* treatment, observers only learn only learn what "a participant of this session decided" after stating their beliefs. In the *Picture* treatment, observers see the *public* decision together with a picture of the decision-maker. We take photographs after the participants are seated in the lab and use those within the session. The photographs are deleted directly after the experiment and are never connected to the participants' names. In the *Picture+Bonus* treatment, we show public decisions together with a picture of the decision-maker (like in *Picture*) and additionally pay a monetary bonus to the decision-maker according to the observer's beliefs. The additional bonus to the decision-maker equals the observer's stated probability (in points) that the decision-maker donated in the *private* decision. Therefore, the size of the bonus payment is always between 0 and 100 points. Through the bonus, we further strengthen the image concern incentive but also change its nature. While in *Anonymous* and *Picture* an altruistic appearance is only beneficial in affective terms, in *Picture+Bonus*, it also increases one's payoff.

Within subjects, we vary the value of p, or how likely a donation is to be enacted, produced, or made – i.e., the *cheapness* of the signal. For the first eight rounds, p is 0.9, overruling each pro-donation decision with 90% probability. In rounds 9 to 16, p changes to 0.1, making the signal less *cheap*, as each donation decision is overruled with only 10% probability. To control for order effects, half the subjects have a reversed order and start with $p = 0.1$ in the first eight rounds. We do not announce the change of p in advance; subjects only know there will be a change in the game after eight rounds, but do not know in advance what this change might relate to.

Matching groups are randomly formed at the beginning of each session and consist of eight participants.[1] During the first 16 rounds, there is no interaction. Since we vary the image concern incentives across sessions, subjects of the same matching group encounter the same image concern treatment. In the observer rounds, however, the observed subjects are always randomly drawn within the matching group. As one participant observes the behavior of others 16 times, observers encounter the same subjects multiple times. While stating the probabilities, subjects do not know for which other participant they are stating beliefs, as their identity (in the treatments with picture) is only revealed on the subsequent screen.

2.2 Procedure

The experiment was programmed with ztree (Fischbacher, 2007), and students of the University of Konstanz were recruited via ORSEE (Greiner, 2015). The decision screens and an English translation of the instructions can be found in the online appendixes B and C. We conducted seven sessions with 165 participants in November and December of 2018. The average age was 22.18 ($sd = 2.59$), and

[1] In one session in the *Anonymous* treatment, 21 instead of 24 participants showed up to the laboratory in time. The software then created three matching groups of seven participants in this session.

37.58% ($sd = 0.49$) of the participants were male.[2] In sessions that included pictures (treatments *Picture* and *Picture+Bonus*), before entering the laboratory, we informed the potential participants that we would take a picture of them. We gave them the option to leave immediately. Only one person chose to leave after this announcement. The announcement clarified that (i) we would take a picture at the beginning, (ii) this picture would be shown together with certain decisions to other participants, (iii) it would always be clear which decisions would be shown to others, and (iv) pictures were never connected to names and deleted immediately after the experiment.

2.3 Hypotheses

Varying the strength of image concerns (between subjects) and cheapness of the signal (within subjects) allows for testing how these changes and the interaction between them influence donation behavior. Additionally, as the second half of the experiment consists of belief elicitation about the donation behavior, we can also test for the influence of our treatment variation on the beliefs. In section 3, we will present the results of the tests for the hypotheses developed in this subsection.

To guide our predictions, we use the signaling model from Bénabou and Tirole (2006) and, as in Grossman (2015), extend it by an exogenous probability that the pro-social act does not take place.

Let us assume each decision-maker has an individual, intrinsic, and unobservable level of altruism $\mu_i \sim U(0,1)$. The distribution of altruism is common knowledge, and μ_i shapes how costly individuals experience the pro-social act to be. The reasoning behind this is that more altruistic individuals derive more utility from helping others. For simplicity, we assume that an individual with altruism level μ_i receives utility $\mu_i b \geq 0$ from creating a benefit b to others. This utility sets off a part of the direct costs ($c > 0$) associated with the pro-social act. The net costs of an (implemented) pro-social action are therefore $c - b\mu_i \geq 0$ and lower for individuals with a higher degree of altruism.[3]

Moreover, decision-makers are image-concerned in the sense that appearing altruistic to others generates positive utility. We denote the strength of image concerns with the parameter $\lambda > 0$, which is, for simplicity, identical for all individuals. The utility for being seen as pro-social can be derived from a valuation for an altruistic image per se and/or future benefits linked to such an image. For our purpose, the exact source is not important (following Andreoni and Bernheim, 2009). In our setting, a decision-maker can choose whether to decide in favor ($a_i = 1$) or against ($a_i = 0$) a costly pro-social act. This decision is public and allows others to update their beliefs about an individual's level of altruism based on the individual's decision (a_i) in a given setting.

We then introduce p as an exogenous probability that depicts how likely a donation is to be enacted, produced, or made. The value of p determines the probability that the pro-social act will not

[2] In the first two sessions, each part lasted for 20 instead of 16 rounds. Since 20 rounds turned out to be too time-consuming, we cut it down to 16 rounds for the remaining sessions.

[3] In principle, $(c - b\mu_i) \leq 0$ may also hold, i.e., a donation is not costly for the individual due to a high level of altruism. The interesting case for us is, however, when the donation is costly, because decisions may change when the image concern or the cheapness changes. Otherwise, one should always donate independently of the specific setting.

be carried out after deciding in favor of it.

In the experiment, the costly pro-social act in which people can engage or not is donating to a charity. Specifically, participants have to decide between donating $100 \cdot X$ points or keeping 100 points for themselves with $X \in \{0.5; 0.75; 1.5; 2; 2.5; 3\}$. We refer to X as the efficiency factor because it shows how efficient it is to give up 100 points. The individual knows the value of X. Moreover, he knows that, with probability p, if he decides in favor of donation, this donation will not take place and he must keep 100 points. Therefore, a donation that is implemented clearly causes direct costs of $c = 100$ points and a benefit of $b = 100 \cdot X$ points for the charity. After the individual decides, his decision and the exact decision setting (p, X) is publicly revealed and allows an observer to draw inferences about the decision-maker's μ_i. This is common knowledge. Adopting all other assumptions from above, with an individual's level of altruism ($\mu_i \in (0, 1)$), the strength of image concerns (λ), and the round budget (100 points) as a reference point and having to decide whether to donate points ($a_i = 1$) or not ($a_i = 0$), the expected utility becomes:

$$EU(\mu_i, a_i, p, X, \lambda) = \underbrace{\lambda E[\mu_i | a_i, p, X, \lambda]}_{\text{image}} - \underbrace{a_i(1 - p)(100 - 100X\mu_i)}_{\text{net costs}} \tag{1}$$

In the experiment, the strength of the image concerns λ varies between treatments. By showing a picture together with the decision in the *Picture* and the *Picture+Bonus* treatment, λ is expected to be higher than in the *Anonymous* treatment.

This expected utility allows calculating the Perfect Bayesian Equilibrium (PBE) of the game. In such a PBE, observers assume an action ($a_i \in \{0, 1\}$) for each type of decision-maker (for each μ_i) and update their belief about the decision-maker's type after observing the action. Taking these beliefs as given, a decision-maker chooses an action that maximizes his expected utility. If the optimal strategies for each type of decision-maker align with the observers' beliefs over the strategies, the conditions for the PBE are fulfilled.

For a continuous distribution of μ_i, a threshold strategy of the decision-maker constitutes such an equilibrium. In such a threshold strategy, decision-makers with $\mu_i \geq \bar{\mu}$ decide in favor of the pro-social act, and decision-makers with $\mu_i < \bar{\mu}$ decide to keep the money. Observing the action then allows inferring whether a decision-maker's level of altruism is above or below the equilibrium threshold. Taking the distribution of μ_i into account, the observer updates $E[\mu_i | a_i = 1] = E[\mu_i | \mu_i \geq \bar{\mu}] = \frac{\bar{\mu}+1}{2}$ and $E[\mu_i | a_i = 0] = E[\mu_i | \mu_i \geq \bar{\mu}] = \frac{\bar{\mu}}{2}$. Following equation (1), the threshold level of altruism that renders the decision maker indifferent between donating and keeping the money can then be calculated as

$$\bar{\mu} \equiv \bar{\mu}(\lambda, X, p) = \frac{1}{X} \cdot \left(1 - \frac{1}{200} \cdot \frac{\lambda}{(1 - p)}\right) \tag{2}$$

In the experiment, we vary p and X (within subjects) and λ (between subjects) to test how such a change affects the donation behavior (a_i) as well as the beliefs of the observer ($E[\mu_i | a_i = 0]$ and

$E[\mu_i|a_i = 1]$).

From (2) we can derive (3),(4), and (5).

$$\frac{\partial \bar{\mu}}{\partial p} = -\frac{\lambda}{200 \cdot X \cdot (1-p)^2} < 0. \tag{3}$$

$$\frac{\partial \bar{\mu}}{\partial X} = -\frac{1}{X^2}\left(1 - \frac{\lambda}{200 \cdot (1-p)}\right) < 0. \tag{4}$$

$$\frac{\partial \bar{\mu}}{\partial \lambda} = -\frac{1}{200 \cdot X \cdot (1-p)} < 0. \tag{5}$$

As already shown by Grossman (2015), the equilibrium threshold of altruism where people switch from keeping the money to donating decreases with the probability that donation cannot occur. Increasing the probability that one will not donate after a decision in favor of donation makes the signaling *cheaper*. Therefore, a larger share of subjects should choose to donate when increasing p. Since we assume uniformly distributed types, the relative image benefit is constant in our model. For other distribution assumptions, however, the relative image benefit may also decrease when p increases. Yet, the prediction that the marginal net image gain from deciding in favor of the pro-social option instead of against the pro-social option is positive holds for such a monotone cutoff equilibrium as long as the decrease in the relative image benefit is lower than the decrease in the expected costs (which should be the case in stable equilibria, see Grossman (2015) and Bénabou and Tirole (2006)). Monotonicity here refers to the idea that higher types have a higher net benefit from the pro-social option than lower types. Further, following intuitive reasoning, donations should occur more often if, ceteris paribus, the donation is more efficient (X) and if individuals care more about their image (λ).

Changing p and λ in the experiment also allows to test for an interaction effect. According to the model, the effect of p on the share of participants choosing to donate increases by increasing the level of image concerns λ:

$$\frac{\partial^2 \bar{\mu}}{\partial p \partial \lambda} = -\frac{1}{200 \cdot X \cdot (1-p)^2} < 0 \tag{6}$$

The intuition behind this prediction is as follows. The strength of image concerns λ defines how much individuals value the relative image benefit of switching to a donation decision,

$$\lambda \left(E[\mu_i|\mu_i \geq \bar{\mu}] - E[\mu_i|\mu_i < \bar{\mu}]\right). \tag{7}$$

If the image concern incentives are high, image gains are more valuable, and we expect that individuals will exploit the chance to improve their image cheaply more often. Consequently, we expect that cheapened signaling costs (due to increases in p) cause more switches to donation decisions when λ is higher.

Derived from the model above, the hypotheses about the donation decisions that can be tested with our experimental setup can be summarized as follows:

Hypothesis 1a. *Public* donation decisions are more frequent ...

- ... if the donation is unlikely ($p = 0.9$) instead of likely ($p = 0.1$) after deciding in favor of it.
- ... if the efficiency factor X increases.
- ... when increasing the level of image concerns. There should be more donation decisions in the *Picture* treatment than in *Anonymous*, and more donation decisions in *Picture+Bonus* than in *Picture*.

Hypothesis 1b. The effect of p on *public* donation decisions

- ... is smaller in the *Anonymous* treatment than in the *Picture* treatment.
- ... is smaller in the *Picture* treatment than in the *Picture+Bonus* treatment.

The literature indicates the relevance of self-signaling motives in pro-social decision settings. Also in our setting, self-signaling motives are expected to play a role in each pro-social decision, the public ones as well as the private ones. In the public decisions, self-signaling motives also predict a larger p to increase the share of decisions in favor of donation. As we cannot rule out self-signaling, we tackle this issue by randomly allocating the participants into the different image concern treatments. In each treatment, the motives to self-signal should be equal, as we vary only the image concerns towards another person between treatments. Any difference in treatments should therefore not be driven by self-signaling (self-image) considerations, but only by social-signaling (social-image) considerations.

The model above also allows for forming predictions on an observer's belief-updating. As an observer's beliefs play an important role in these kinds of model (they determine the incentives for the decision-maker), we consider it crucial to test whether the observer's belief-updating follows the model's predictions. If it becomes cheaper to signal altruism (p increases), observers should discount the signaling value of pro-social decisions, i.e., deciding for donation becomes a *worse* signal. However, keeping the 100 points also becomes a *worse* signal because the observed person does not decide pro-socially, even though it would have been cheaper to do so. Similarly, if image concerns become more important, a larger share of decision-makers is predicted to decide in favor of the donation. Observing a decision in favor of donation for higher λ should therefore lead to a lower updated belief about the level of altruism. Similarly, observing a decision to keep the money in the case of a higher image concern should also decrease the believed level of altruism of the decision-maker compared to the case of lower image concerns. For the decision-maker's trade-off between the two options, the observer's updated beliefs about the decision-maker's level of altruism after both options are important. We will therefore test the belief-updating after a donation decision ($a_i = 1$), as well as after a keep decision ($a_i = 0$).

Hypothesis 2a. The perceived degree of altruism following a donation decision ($a_i = 1$) **and** following a keep decision ($a_i = 0$) is lower...

- ... the less likely a donation is to be enacted, i.e. the higher p.

- ... the higher the image concerns (λ) of the decision-maker.

When individuals value the relative image benefit of switching to a donation decision more with a higher level of λ, individuals have higher incentives to exploit the chance to improve their image cheaply (hypothesis 1b). Consequently, by increasing the image concerns, when the share of individuals who choose to donate changes more (due to a change in p), the observer's beliefs should also change more. The same argument applies to keeping decisions, as a smaller share of people is predicted to keep the points with a higher image concern. This share is predicted to decrease even more when changing p. The change in equilibrium beliefs for this pattern is outlined in hypothesis 2b.

Hypothesis 2b. The effect of p on the perceived degree of altruism following a donation decision ($a_i = 1$) **and** following a keep decision ($a_i = 0$)...

- ... is smaller in the *Anonymous* treatment than in the *Picture* treatment.
- ... is smaller in the *Picture* treatment than in the *Picture+Bonus* treatment.

3 Results

3.1 Decision-making behavior

Hypotheses 1a and 1b make specific predictions about how subjects change their donation behavior when signaling altruism gets cheaper and when image concerns increase, and how these two effects should interact. We now present how the subjects actually behaved. In hypothesis 1a, we surmised that subjects would exploit the opportunity to cheaply signal altruism and decide more often for donation than they would otherwise do. That is, we expected that more subjects would publicly indicate their willingness to donate if the donation is unlikely ($p = 0.9$). Columns (1) and (2) in table 1 show that overall, the first part of hypothesis 1a can be confirmed. Taking all decision-data into account shows a significant increase in donation decisions if donation becomes more unlikely. Similarly, the same regressions also show that the second part of hypothesis 1a can be confirmed: higher X leads to more donation decisions.
Figure 1 plots the share of donation decisions for each treatment and each efficiency factor, as well as for both values of $p \in \{0.1, 0.9\}$. Panel 1a suggests that participants whose donation decisions were disclosed anonymously indeed reacted to cheapened signaling costs. For each value of the efficiency factor, we always observed more donation decisions if deciding for donation became cheaper ($p = 0.9$ instead of $p = 0.1$). On average, cheapening signaling costs increased the percentage of donation decisions by 8 percentage points (p-value = 0.008, table 1). Figure 1b shows a similar pattern for the *Picture* treatment. In the *Picture* treatment observers saw the donation decisions together with a photo of the decision-maker. Figure 1b shows the influence of p on donation behavior is weaker in the *Public* than in the *Anonymous* treatment and, as can be seen in table 1, also not statistically significant.

However, clearly different is the *Picture+Bonus* treatment. In figure 1c, it is straightforward to see that behavior does not change systematically if the donation becomes less likely from $p = 0.1$ to

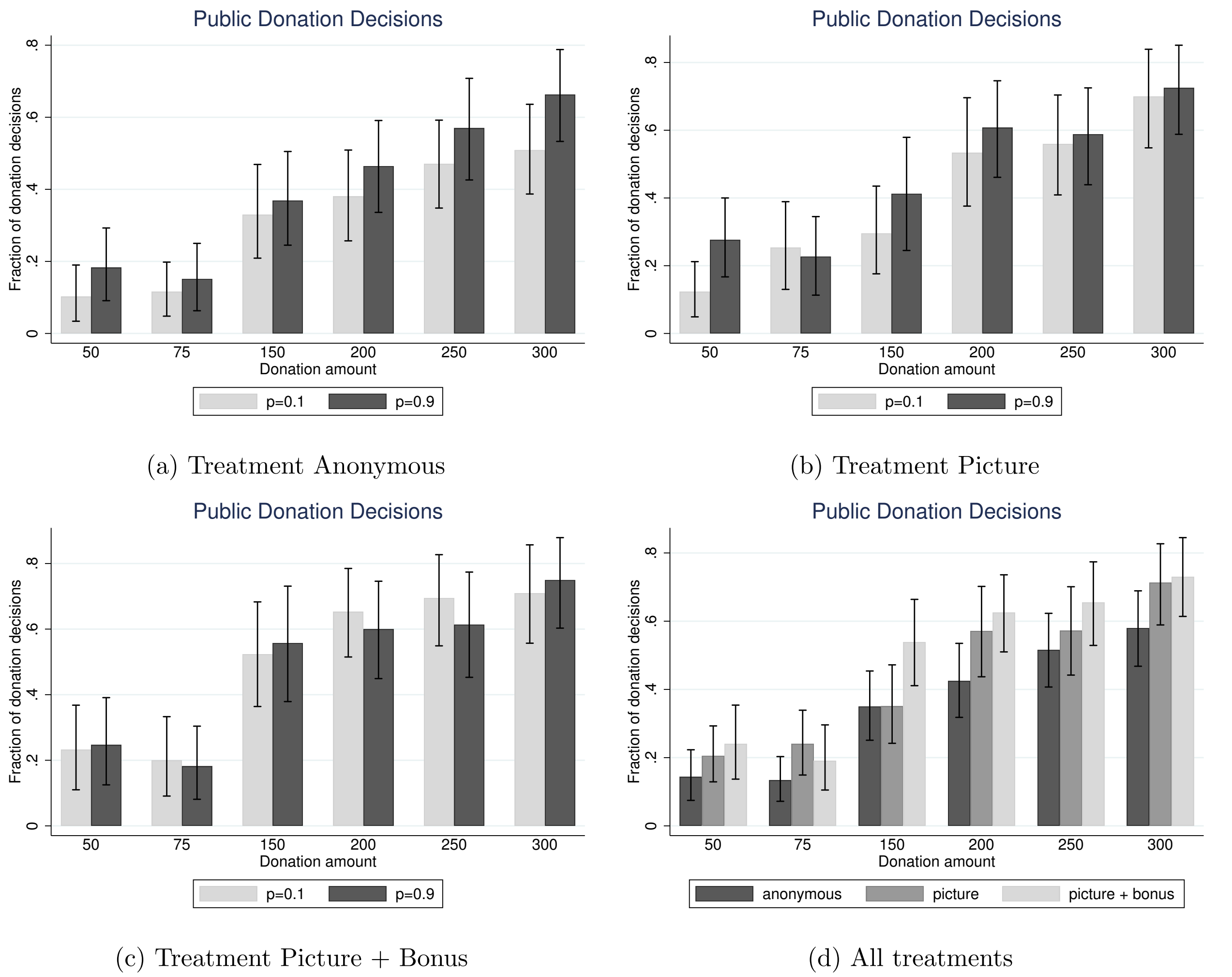


Figure 1: Share of public donation decisions.

Notes: Whiskers represent the 95% confidence intervals. Standard-errors are bootstrap with 10,000 repetitions and clustering at subject level.

$p = 0.9$. Correspondingly, cheapened signaling costs do not have a significant effect on the donation decisions in the regression analysis (table 1, columns (7) and (8)). Seemingly, paying a monetary bonus for a pro-social appearance additionally prevents an effect from occurring. A crowding-out effect can potentially explain this difference between extrinsic and intrinsic motivation (as modeled in Bénabou and Tirole, 2006) and as also shown in different settings, e.g. in Frey and Oberholzer-Gee (1997); Gneezy and Rustichini (2000); Newman and Shen (2012); Müller and Rau (2020).

Table 2 columns (1) and (2) show what can also be observed in figure 1d. The share of donations seems to increase by increasing the image concerns from *Anonymous* to *Picture*. Comparing the *Anonymous* and *Picture* treatment, the difference in donations is 8.6 percentage points ($p = 0.077$). Comparing the *Picture* and *Picture+Bonus* treatments, the difference in donations is only 5.6 percentage points and not significantly different from zero. All subfigures in figure 1, and all regressions in table 1 and table 2, columns (1) and (2), further confirm that the efficiency factor in the public

Table 1: Public donation decisions

	All data		Anonymous		Picture		Picture+bonus	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Donation unlikely	.052*	0.052*	0.083**	0.083**	0.065	0.065	-0.012	-0.013
($p = 0.9$)	(.021)	(0.022)	(0.030)	(0.031)	(0.039)	(0.040)	(0.041)	(0.043)
Efficiency Factor	.202***	.204***	0.191***	0.195***	0.206***	0.205***	0.219***	0.217***
(X)	(.018)	(0.018)	(0.026)	(0.026)	(0.032)	(0.033)	(0.035)	(0.036)
cons.	.052		-0.009		0.059		0.132	
	(.032)		(0.042)		(0.053)		(0.073)	
Obs.	2832	2832	1200	1200	864	864	768	768
Clusters	165 subj.	165 subj.	69 subj.	69 subj.	48 subj.	48 subj.	48 subj.	48 subj.
Subject F.E.	no	yes	no	yes	no	yes	no	yes
R^2	0.134	0.421	0.126	0.464	0.141	0.362	0.156	0.402

Notes: OLS regressions on deciding in favor of donation in the *public* situation. Columns (3) and (4) contain data of treatment *Anonymous*. Columns (5) and (6) contain data of treatment *Picture*. Columns (7) and (8) contain data of treatment *Picture+Bonus*. Columns (2), (4), (6) and (8) include subject fixed effects. Conducting the regressions of table 1 with logit instead of OLS does not change the results qualitatively.
Std. errors are clustered at the subject level and depicted in parentheses. ***(**/*) significant at the 0.1 (1/5) percent level.

donation decision increases the share of subjects choosing donation highly significantly.

Table 2 addresses hypothesis 1b and reports the interaction results between increasing the cheapness of the signal and the level of image concerns.[4] The model in section 2.3 predicts that the change in p influences the share of donation decisions more the higher the image concerns are. The interaction terms of the different treatments and p in table 2 column (2) reveal that the data cannot reject the hypothesis that there is no difference. Omitting the treatment *Picture* in the regressions, and showing the additional effect of p in the *Anonymous* treatment and the *Picture+Bonus* treatment, shows no significant coefficients. The results in columns (3)-(8) show that running the regressions separately for different bins of X does not change this result qualitatively.

To see whether learning or time effect might play a role (as in Arechar and Rand (2022)), we provide the average donation shares over the different periods in the appendix figure A2a. The figure shows that the time within the experiment does not seem to change behavior systematically. The second donation decision in each round is private and mainly used to incentivize the observer's beliefs about an unobservable action of the decision-maker. We summarize the behavior in the private situation briefly in subsection 3.3.

3.2 Belief-updating

Cheapening signaling costs should also influence how others evaluate donation decisions. After all, donation decisions that are implemented with a low probability should raise doubts about whether high altruism (instead of the cheapness of the situation) was the dominant motive behind the deci-

[4]Repeating the regressions with logit instead of OLS does not change the results qualitatively.

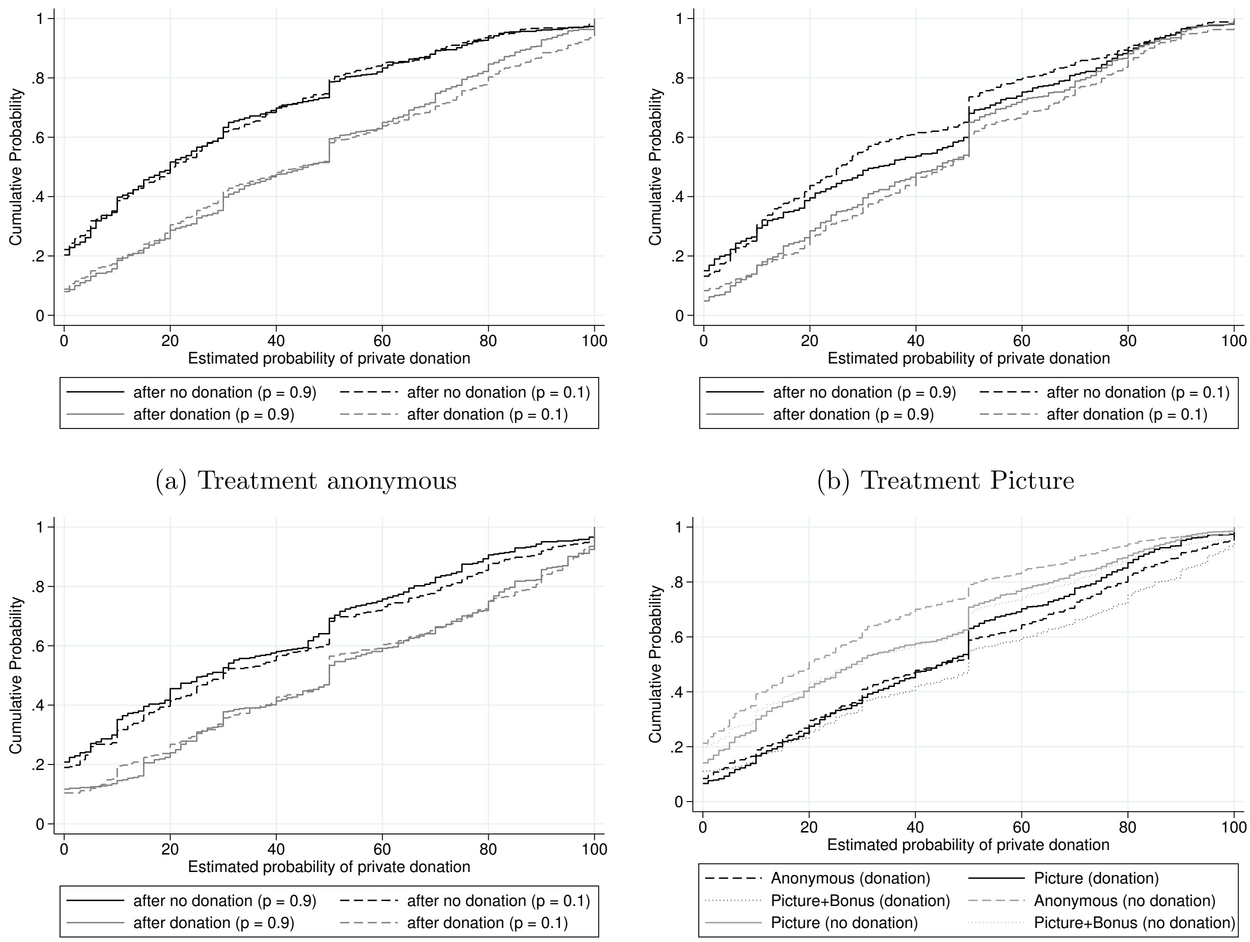
Table 2: Choosing donation in the public decision

	All Data		$X < 1$		$1 < X \leq 2$		$2 < X \leq 3$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Efficiency Factor (X)	0.203***	0.203***						
	(0.018)	(0.018)						
Donation unlikely (p = 0.9)	0.052*	0.065	0.043	0.064	0.056	0.103	0.061	0.039
	(0.021)	(0.039)	(0.026)	(0.047)	(0.037)	(0.061)	(0.031)	(0.056)
Treatment: Anonymous	-0.086	-0.095	-0.084	-0.080	-0.079	-0.060	-0.092	-0.133
	(0.048)	(0.054)	(0.052)	(0.055)	(0.071)	(0.083)	(0.074)	(0.084)
Treatment: Picture+Bonus	0.056	0.095	-0.003	0.029	0.120	0.175*	0.051	0.079
	(0.049)	(0.058)	(0.063)	(0.071)	(0.071)	(0.088)	(0.075)	(0.088)
Treatment: Anonymous		0.019		-0.007		-0.039		0.089
× donation unlikely (p = 0.9)		(0.049)		(0.061)		(0.081)		(0.073)
Treatment: Picture+Bonus		-0.077		-0.062		-0.113		-0.055
× donation unlikely (p = 0.9)		(0.057)		(0.070)		(0.102)		(0.082)
cons.	0.071	0.064	0.200***	0.189***	0.441***	0.418***	0.612***	0.623
	(0.042)	(0.045)	(0.042)	(0.046)	(0.056)	(0.063)	(0.057)	(0.064)
Obs.	2832	2832	913	913	987	987	932	932
Clusters	165 subj.	165 subj.	165 subj.	165 subj.	165 subj.	165 subj.	165 subj.	165 subj.
R^2	0.148	0.150	0.014	0.015	0.029	0.031	0.020	0.023

Notes: OLS regressions on deciding in favor of donation. Columns (1) and (2) contain data of all public decisions (Treatment *Picture* is omitted). Columns (3) and (4) contain data with efficiency factor $X \in \{0.5; 0.75\}$. Columns (5) and (6) contain data with efficiency factor $X \in \{1.5; 2\}$. Columns (7) and (8) contain data with efficiency factor $X \in \{2.5; 3\}$. Std. errors are clustered at the subject level and depicted in parentheses. ***(**/*) significant at the 0.1 (1/5) percent level.

sion. Observers should understand that one does not have to be extraordinarily altruistic to commit to a donation that will most likely not be implemented. Consequently, the expected altruism and the estimates should be lower when donations are less likely (see the first part of hypothesis 2a). Similarly, higher image concerns during the public donation decisions should also decrease the observer's estimates. If the image concerns for a donation is higher, one does not need to be extraordinarily altruistic to choose to donate (see the second part of hypothesis 2a). We will now test whether observers' estimates follow this principle.

Recall that observers' estimates are the incentivized belief that a person donated in the private decision given their public decision. Observers were first informed about (i) how likely donation was for the observed person in the public decision and (ii) how much the observed person could donate in private and public if the donations were carried out. They then had to make two estimates for the likelihood of donating in the private decision: one assuming that the observed person had decided to donate and one assuming that the observed person had decided to keep the 100 points in the public decision. Our quadratic scoring rule for binary outcomes maximized the observer's expected payoff by reporting the true beliefs about the likelihood of donation.



(a) Treatment anonymous

(b) Treatment Picture

(c) Treatment Picture + Bonus

(d) All treatments

Figure 2: c.d.f. of observers' beliefs about a private donation in the different treatments.

Notes: Beliefs following a public no-donation decision are in black, and a public donation decision are in gray in panels (a)-(c). Solid lines in panels (a)-(c) represent the *cheaper* signals, as 90% of the public donation decisions will not be carried out. Panel (d) plots the beliefs in the different treatments, not separated by different values of p.

Table 3: Observers estimate about likelihood of private donation

	Conditional on public donation						Conditional on no public donation					
	All Data			Anonym.	Pic.	Pic + Bon.	All Data			Anonym.	Pic.	Pic + Bon.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Efficiency Factor	-2.126*	-2.051*	-2.053*	-2.268	-0.007	-3.970*	-2.034*	-1.993*	-1.993*	-1.624	-0.597	-4.117*
(Public Donation)	(1.020)	(1.001)	(1.003)	(1.172)	(1.791)	(1.985)	(0.979)	(0.975)	(0.981)	(1.218)	(1.978)	(1.640)
Efficiency Factor	8.02***	8.027***	8.027***	8.336***	8.687***	6.883***	5.525***	5.495***	5.496***	4.786***	5.022*	7.112***
(Private Donation)	(1.193)	(1.191)	(1.197)	(1.991)	(2.296)	(1.615)	(1.153)	(1.122)	(1.108)	(1.239)	(2.542)	(1.796)
Public donation unlikely	-0.819	-0.817	-1.909	-0.939	-1.838	0.509	0.416	0.417	3.509*	0.094	3.519**	-2.491
($p = 0.9$)	(0.713)	(0.713)	(1.257)	(1.243)	(1.123)	(1.097)	(1.094)	(1.094)	(1.392)	(1.632)	(1.280)	(2.129)
Treatment: Anonymous		1.734	1.244					-6.957	-5.241			
		(2.964)	(3.076)					(3.593)	(4.032)			
Treatment: Picture+Bonus		6.122	4.874					0.419	3.439			
		(3.323)	(3.258)					(2.765)	(3.085)			
Treatment: Anonymous			0.980						-3.432			
$\times$ *donation unlikely* ($p = 0.9$)			(1.852)						(2.187)			
Treatment: Picture+Bonus			2.495						-6.040*			
$\times$ *donation unlikely* ($p = 0.9$)			(1.658)						(2.657)			
cons.	36.333***	33.811***	34.359***	35.450***	29.603***	44.490***	26.320***	29.134***	27.588***	22.920***	25.966***	31.780***
	(3.039)	(2.916)	(2.904)	(3.893)	(2.159)	(7.198)	(3.164)	(2.449)	(2.512)	(4.649)	(4.860)	(5.552)
Obs.	2832	2832	2832	1200	864	768	2832	2832	2832	1200	864	768
Clusters	21 MG	21 MG	21 MG	9 MG	6 MG	6 MG	21 MG	21 MG	21 MG	9 MG	6 MG	6 MG
R^2	0.057	0.063	0.063	0.535	0.369	0.534	0.031	0.045	0.047	0.572	0.468	0.527

Notes: OLS regressions on observers' stated probability that decision-maker will donate in the private situation (conditional on decision in the public situation). Columns (1)-(6) report regressions of the observers' beliefs after a public donation decision, columns (7)-(12) after a public no-donation decision. Columns (4)-(6) and (10)-(12) only contain data of one image concern treatment each. Clustered standard errors are bootstrapped with 10 000 replications and depicted in parentheses. ***(**/*) significant at the 0.1 (1/5) percent level.

Figure 2, panels (a)-(c), plot the cumulative distribution functions of the observers' estimates for each treatment. The figures show that the belief about a donation in the private situation after a public donation decision is distributed to the right (higher probabilities) of the beliefs about a donation in the private situation after a public decision against donation. Further, the figure does not show large differences between the beliefs in rounds where the public donation was *cheap* ($p = 0.9$) and the beliefs in rounds where the public donation was less cheap ($p = 0.1$). The model would predict the dashed lines to be to the right of the respective solid line after the same public decision. After a donation decision, the decision should be a weaker positive signal in the case of $p = 0.9$, and the beliefs for the private donation decision should therefore decrease. Similarly, after a no-donation decision, the decision should be a stronger negative signal in the case of $p = 0.9$, and the belief about the private donation decision should decrease too. These patterns cannot be observed in figure 2. This suggests that the null hypothesis associated with first part of hypothesis 2a cannot be rejected: we cannot show that a higher p decreases observers' beliefs.

Figure 2, panel (d), contains the cdf of observers' beliefs for all three image-concern treatments. The second part of hypothesis 2a predicts a lower belief for higher image concerns, both after a donation decision and after a no-donation decision. The cdfs after a donation decision for the *Picture* and *Picture+Bonus* treatment look very similar. Further, after a donation decision, the ordering of the cdf does not follow the strength of the image concerns. This suggests that the null hypothesis associated with the second part of hypothesis 2a can also not be rejected: we do not see that higher image concerns systematically decrease observers' beliefs.

However, the graphs in figure 2 may be somewhat unbalanced, as they do not differentiate between the different efficiency factors that were randomly and independently drawn for each *public* and *private* situation. Table 3 shows how the efficiency factor in the public decision, the efficiency factor in the private decision, and the likelihood of donation in the public decision influence observers' estimates after a public donation (columns (1)-(6)) or a public keep-decision (columns (7)-(12)). To unravel potential differences across treatments, columns (4)-(6) as well as columns (10)-(12), treat each image concern treatment as a separate unit. As observers interact within the matching groups, we cluster at the matching group level. To avoid inference issues due to the low number of matching groups, we report bootstrapped clustered standard errors in table 3.

Columns (1)-(3) of table 3 report the regression results of the observers' estimates after a donation decision in the public situation for all three treatments together. A higher efficiency factor in the private donation significantly increases the observer's belief that the observed subject donates in the private decision. A higher efficiency factor in the public decision decreases the estimate. This is in line with the intuition that public donations with a high efficiency are a weaker signal for a decision-maker's level of altruism than donation decisions with a low efficiency. Column (2) reveals that the beliefs after a public donation decision do not significantly differ between the *Picture* treatment and the other treatments. The table further shows that p does not significantly influence the observers' beliefs. This observation that there is no significant correlation between the value of p and the observer's belief after a public donation decision also holds for each of the three treatments separately (columns (4)-(6)). The model would predict a systematic negative relationship here, as a decision that is less likely to be relevant should (ceteris paribus) also decrease the informational content a donation decision carries about the level of altruism and therefore decrease the observer's stated belief about the donation in the private situation. We therefore fail to reject the null-hypothesis associated with both parts of hypothesis 2a for the beliefs following a donation decision. Hypothesis 2b predicts that the effect of p will be larger for treatments with higher image concerns. Column (3) does not show any significant differences between the impact of p on the estimates in the *Picture* treatment and the other treatments. We cannot, therefore, for either part of hypothesis 2b, reject the associated null hypothesis of no difference in beliefs between the different levels of image concerns after a donation decision.

Columns (7)-(12) of table 3 report the results of regressing the observer's beliefs about the likelihood of a donation in the private setting after a no-donation decision in the public setting. A higher efficiency factor in the public decision decreases the estimate. This is in line with the intuition that public decisions against a donation with high efficiency are a stronger signal for a decision-maker's low level of altruism than public decisions against a donation with low efficiency. The higher efficiency factor in the private situation significantly increases the observer's estimates. Columns (7) and (8) show that, overall, neither the *cheapness* of the public signal nor the difference in image concerns between *Picture* and *Anonymous*, or *Picture* and *Picture+Bonus*, leads to a significant difference in observers' estimates. Columns (10)-(12) also show that there is no significant negative effect of p on the observer's estimate after a no-donation decision in any of the three image concern treatments. The model predicts a negative coefficient here: not choosing to donate even though it would be very *cheap* to do so (the donation would anyhow not be carried out with 90% probability) should be a strong signal of a participant's selfishness. Following this logic, the observer should enter a lower estimate about the probability that the participant will donate in the private situation. This prediction is not visible in the data. Column (11) even shows that, in the *Picture* treatment, a higher p leads to significantly *higher* estimates following a no-donation decision. Further, as hypothesis 2b states, the (negative) effect of increasing p on the estimates should be stronger when increasing the image concern. However, column (9) shows that we fail to reject the null hypothesis of no difference, even though, compared to the increase in beliefs for higher p in the *Picture* treatment, the additional effect of a high p in the *Picture+Bonus* treatment is significantly negative. Comparing the effect of p on the estimates between the *Picture* and the *Anonymous* treatment does not show up as significant. Therefore, we conclude that the null hypotheses of no difference associated with hypotheses 2b also cannot be rejected.

After observers made their estimates for both possible cases – one for the scenario that the observed person donated in the public situation and one for the scenario that the observed person did not donate in the public situation – they learned how the observed person had decided in the public situation. Potentially, observers might use this feedback about the public behavior of others to update their estimates about donations in the private situation. We plotted observers' estimates across rounds to see whether observers changed their estimates over time. Figures A3 and A4 in the appendix do not suggest that this is the case: estimates appear to have been relatively constant over time. Thus, there seems to be no apparent trend in updating or learning behavior.

3.3 Private donations

Our experiment is set up to measure the level of altruism of decision-makers and to measure the beliefs about this level of altruism by an observer. We vary a public donation situation to determine the influencing factors of an observer's belief about the decision-makers underlying level of altruism. To incentivize the accuracy of beliefs about altruism, we let decision-makers choose again about a donation in each round. This second donation decision in each round is the same throughout the treatments and only differs by the efficiency factor. There is no factor p that influences the likelihood of donation, and no difference in the level of image concern, as all these decisions are done in private and not shared with any other participant. Self-signaling or self-image concerns might of course also be apparent in private donation decisions. The absolute level of the beliefs can mix the effect of true altruism and self-signaling. This driving force of donations, however, should not vary between treatments. Our experiment, therefore, has the potential to cleanly observe how the different situations in public decisions change beliefs about altruism.

To examine the validity of our measure of altruism, we check whether subjects donating in the public situation are also more likely to donate in the private situation and whether the private donation decisions do not vary systematically between the treatments. The regression analyses in table A1 in the appendix confirm the measure's validity. Columns (1) and (2) show that donation decisions in the public situation correlate significantly with donation decisions in the private situation. Further, columns (3)-(5) show that none of the treatment differences for the public situation (the cheapness of the signal, the level of image concern, and the interaction between them) significantly correlate with the donation decisions in the private situation.

Figure A1 in the appendix further shows that, in line with the results in table A1, private donations increase with the efficiency factor of the donation. The figure also shows that the private donations are, in general, at a lower level than public donations for the same efficiency factor. Further, as with the previous results, the private donations do not follow any systematic time trend within the experiment (figure A2b).

3.4 Effect sizes and sample sizes

To assess the robustness of the main results, we now discuss the sample size, power and minimum detectable effect sizes (MDEs) for hypotheses 1a and 2a. Following Vasilaky and Brock (2020), we first calculate the MDEs given the sample sizes and standard deviations in the data, as well as the standard values of $\alpha = 0.05$ and $\beta = 0.8$ for significance and power. We compare these MDEs with the effects we found in the data. The size of the MDEs helps us judge how far some of the null results are likely to be a consequence of a small sample.

Table A2 in the appendix lists, for the different statements in hypotheses 1a and 2a, (i) the MDE given the number of observations and the standard deviation of the outcome variable in the data (but ignoring the cluster dependencies of the data), (ii) the MDEs given the number of clusters and observations in each cluster, the standard deviation in the data and the intra-cluster correlation in the data, and (iii) the effect size observed for each mean comparison in the data. To keep these ex-post power considerations as straightforward as possible, we conduct these calculations for a standard t-test for each hypothesis. As hypotheses 1a and 2a are without interactions, standard one-sided t-tests for binary outcomes can be applied.[5]

Table A2 reveals that, without clustering, an effect of 4.6 to 6.2 percentage points can be detected for the donation decisions at a significance level of 0.05 with a power of 0.8. For the observers' beliefs, differences of between 2.8 and 3.8 can be detected at a significance level of 0.05 with a power of 0.8 with the given sample size when ignoring the clustered structure of the data. These effect sizes do not seem out of reach, as the difference in donations between low and high efficiency factor, for example, is many times higher, and the difference between the *Picture* and *Anonymous* treatment is also higher than this MDE. The numbers for the observers' beliefs also do not seem impossible to find in such an experiment, as the difference in the beliefs between the *Picture* and *Anonymous* treatment, for example, is around 7 percentage points. For the beliefs, however, we can see that the effect in the data is sometimes opposite to what is predicted.

When accounting for the conservative clustering in the tests (individuals for donation decisions and matching groups for observers' beliefs), however, we can see that the calculated MDE to reach a power of 0.8 is quite high with values between 4.1 and 12.6. Ex-post, such effect sizes seem overly optimistic and therefore hard to detect with the approach of clustering the standard errors of the tests as conservatively as possible. A future replication with more data, especially to have sufficient power to test the interaction effects of hypotheses 1b and 2b reliably, might therefore be reasonable when clustering the standard errors on the matching group level.

4 Discussion and conclusion

We study pro-social decision-making in an environment where decisions can serve as costly signals of an individual's level of generosity. Motivated by real-world environments, we focus on decisions that may or may not be carried out. In theory, decreasing the probability that actions have to be carried out after deciding pro-socially reduces the threshold of altruism needed to decide pro-socially. In signaling models with a pro-social context, decision-makers trade off the benefits of image gain with the costs associated with the pro-social action. In equilibrium, beliefs should adapt to the new situation and not influence the marginal image change between deciding in favor of or against the pro-social option. According to the model, as the updated beliefs in a PBE after a pro-social decision *and* after a non-pro-social decision decrease, a lower pivotality of the decision can only increase the share of people deciding for the pro-social option. We find evidence of this in the treatment with

[5] For the efficiency hypotheses we binarized X in two bins: $X_{low} \in \{0.5, 0.75, 1.5\}$ and $X_{high} \in \{2, 2.5, 3\}$.

existing but low image concerns. However, when increasing the level of image concern, the effect of the reduced pivotality on decision-making decreases – which is opposite to the model's prediction. As previous literature finds inconsistent evidence of the effect of cheaper signals on decision-making in a pro-social context, we try to understand the model's flaws better. Since beliefs and the updating of observers determine the strength of image concerns in such contexts, we also measure belief-updating in an incentive-compatible way. This exercise reveals that belief-updating does not seem to be in line with the model's predictions. We cannot reject the null hypothesis that beliefs about the level of altruism do not decrease when making signals cheaper. For beliefs after a decision against the pro-social action, increasing the cheapness of the signal in fact significantly increased the beliefs about the level of altruism of the decision-maker in one of our treatments.
Using the data from our experiment, we cannot rule out that the lack of a significant change in updating for cheaper signals is an artifact of our experimental design and potentially different from other signaling settings. A larger sample size may allow us to find significant effects of the change in probability on belief-updating. However, in the *Picture* treatment, we find a significant positive effect of an increase in cheapness on the beliefs, and our power calculation shows that a small sample size is unlikely to be the reason for this non-finding. This mismatch of predicted behavior derived by the PBE and observed behavior suggests that the model does not capture the relevant driving forces. The model that assumes decision-makers trade off the image gain with the expected net costs of the pro-social act cannot explain this behavior if the image gain does not increase with the net costs of the pro-social behavior. Suppose the self-signaling incentives are sufficiently strong compared to the social signaling incentives. In that case, it can still constitute an equilibrium behavior in that decisions react to p even if the beliefs of observers do not. Even though we tried to make the image benefits towards the one observer somewhat strong (by showing a picture), this one observer's belief may not be sufficient to be taken into account by the decision-maker. Future tests could further increase the social image concerns, for example, by having a larger audience than just one observer. If the observers are again not reacting to the change in pivotality of the decision when forming their beliefs, but the decision-makers are still reacting to the pivotality with their donation behavior, this might be a strong sign that large self-signaling incentives are at play. Another reason for the puzzling observation that decision-makers seem to react to p even if observers do not might be the lack of learning opportunities in our experimental design. As all the donation decisions are made before the beliefs are elicited, decision-makers cannot react to the observers' out-of-equilibrium beliefs. In future studies that aim to focus on the stability of equilibria in such a setting, it might be reasonable to provide feedback on the beliefs to the decision-makers in between decisions. If the beliefs do not react to the pivotality of the decision, we would then also expect the decisions to react less and less over time.

Transferring the results to situations outside of the laboratory implies that clarifying the decision situation's cheapness might help promote pro-social decision-making. In many cheap signaling scenarios, people only have vague ideas about how likely they are to have to fulfill and suffer the costs of their promise. If there is a reason to believe the perceived likelihood is biased upwards and image concerns are low, clarifying how unlikely the costs are to be realized could foster pro-social decisions. We can apply our findings to the examples of the introduction. For organizations that aim to recruit stem-cell donors, it might help to clarify that even after being tested, the possibility of being a match for someone is very low. According to Gragert et al. (2014), the likelihood of finding a match for one of the few thousand people needing a donation is only 75 % in the ethnic group with the highest match likelihood - even though there are more than ten million potential donors in the database. Communication between the organization and the individual is rather anonymous, which might mirror our anonymous setting. Therefore, making the low probability salient could increase take-up and get more people into the database. A journal editor who does not know the potential referees too well could include a sentence to the invitation like *"During the referee period, I might inform you that your report is no longer needed"* to increase the acceptance rate of reviewers. However, reducing the level of anonymity in the experiment reveals that this approach might not improve favorable decision-making in settings with low anonymity. Suppose you ask a colleague you know well in the town of an upcoming conference for their hospitality. In that case, according to our data, the colleague's decision will not necessarily depend on the likelihood of your paper being accepted and that you will travel there. The results of the beliefs are also interesting to transfer into real-world examples. If the behavior in our laboratory experiment is transferable into real-world settings, a person who rejects acting pro-socially, even if it would be very cheap to agree, does not necessarily lead others to think they are very self-centered. The likelihood of the decision eventually being carried out does not seem to be considered as much as theoretical models would predict.

For future research, the robustness of these findings and the potential origins of the deviations between beliefs and theory could be explained further. If beliefs in other signaling settings also turn out to be biased in the case of changing the pivotality, a proper model of equilibrium behavior should account for this. A more precise understanding of the formation of beliefs about pro-sociality, and consequently image concerns as a key ingredient for motivating pro-social behavior, might also be exploited by organizations and allow *nudging* individuals into deciding pro-socially more often.

References

Andreoni, J. and D. Bernheim (2009). Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. Econometrica 77(5), 1607–1636.

Arechar, A. A. and D. G. Rand (2022). Learning to be selfish? a large-scale longitudinal analysis of dictator games played on amazon mechanical turk. Journal of Economic Psychology 90, 102490.

Bénabou, R., A. Falk, and L. Henkel (2022). Ends versus means: Kantians, utilitarians and moral decisions.

Bénabou, R., A. Falk, L. Henkel, and J. Tirole (2022). Eliciting Moral Preferences: Theory and Experiment.

Bénabou, R. and J. Tirole (2006). Incentives and Prosocial Behavior. American Economic Review 96(5), 1652–1678.

Burdea, V. and J. Woon (2022). Online belief elicitation methods. Journal of Economic Psychology 90, 102496.

Carter, J. R. and S. D. Guerette (1992). An experimental study of expressive voting. Public Choice 73(3), 251–260.

Ellingsen, T. and M. Johannesson (2011). Conspicuous generosity. Journal of Public Economics 95(9-10), 1131–1143.

Erkal, N., L. Gangadharan, and B. H. Koh (2020). Replication: Belief elicitation with quadratic and binarized scoring rules. Journal of Economic Psychology 81, 102315.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. Review of Economic Studies 83(2), 587–628.

Falk, A., T. Neuber, and N. Szech (2020). Diffusion of being pivotal and immoral outcomes. Review of Economic Studies 87(5), 2205–2229.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10(2), 171–178.

Fischer, A. J. (1996). A further experimental study of expressive voting. Public Choice 88(1-2), 171–184.

Folli, D. and I. Wolff (2022). Biases in belief reports. Journal of Economic Psychology 88, 102458.

Frey, B. S. and F. Oberholzer-Gee (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. American Economic Review 87(4), 746–755.

Ginzburg, B., J.-A. Guerra, and W. N. Lekfuangfu (2022). Counting on my vote not counting: Expressive voting in committees. Journal of Public Economics 205, 104555.

Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. The Quarterly Journal of Economics 115(3), 791–810.

Gragert, L., M. Eapen, E. Williams, J. Freeman, S. Spellman, R. Baitty, R. Hartzman, J. D. Rizzo, M. Horowitz, D. Confer, et al. (2014). Hla match likelihoods for hematopoietic stem-cell grafts in the us registry. New England Journal of Medicine 371(4), 339–348.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. Journal of the Economic Science Association 1(1), 114–125.

Grossman, Z. (2015). Self-signaling and social-signaling in giving. Journal of Economic Behavior & Organization 117, 26–39.

Klinowski, D. (2021). Reluctant donors and their reactions to social information. Experimental Economics 24(2), 515–535.

Meyer, C. J. and E. Tripodi (2021). Image concerns in pledges to give blood: Evidence from a field experiment. Journal of Economic Psychology 87, 102434.

Müller, S. and H. A. Rau (2020). Motivational crowding out effects in charitable giving: Experimental evidence. Journal of Economic Psychology 76, 102210.

Newman, G. E. and Y. J. Shen (2012). The counterintuitive effects of thank-you gifts on charitable giving. Journal of Economic Psychology 33(5), 973–983.

Nyarko, Y. and A. Schotter (2002). An experimental study of belief learning using elicited beliefs. Econometrica 70(3), 971–1005.

Tonin, M. and M. Vlassopoulos (2013). Experimental evidence of self-image concerns as motivation for giving. Journal of Economic Behavior and Organization 90, 19–27.

Tyran, J. R. (2004). Voting when money and morals conflict: An experimental test of expressive voting. Journal of Public Economics 88(7-8), 1645–1664.

Tyran, J.-R. and A. K. Wagner (2019). Experimental Evidence on Expressive Voting. The Oxford Handbook of Public Choice, Volume 2, 927–940.

Vasilaky, K. N. and J. M. Brock (2020). Power (ful) guidelines for experimental economists. Journal of the Economic Science Association 6(2), 189–212.

Cheap Signaling of Altruism - Online Appendix

Appendix of the paper 'Cheap Signaling of Altruism' by Moritz Janas and Michelle Jordan.

A Further Results

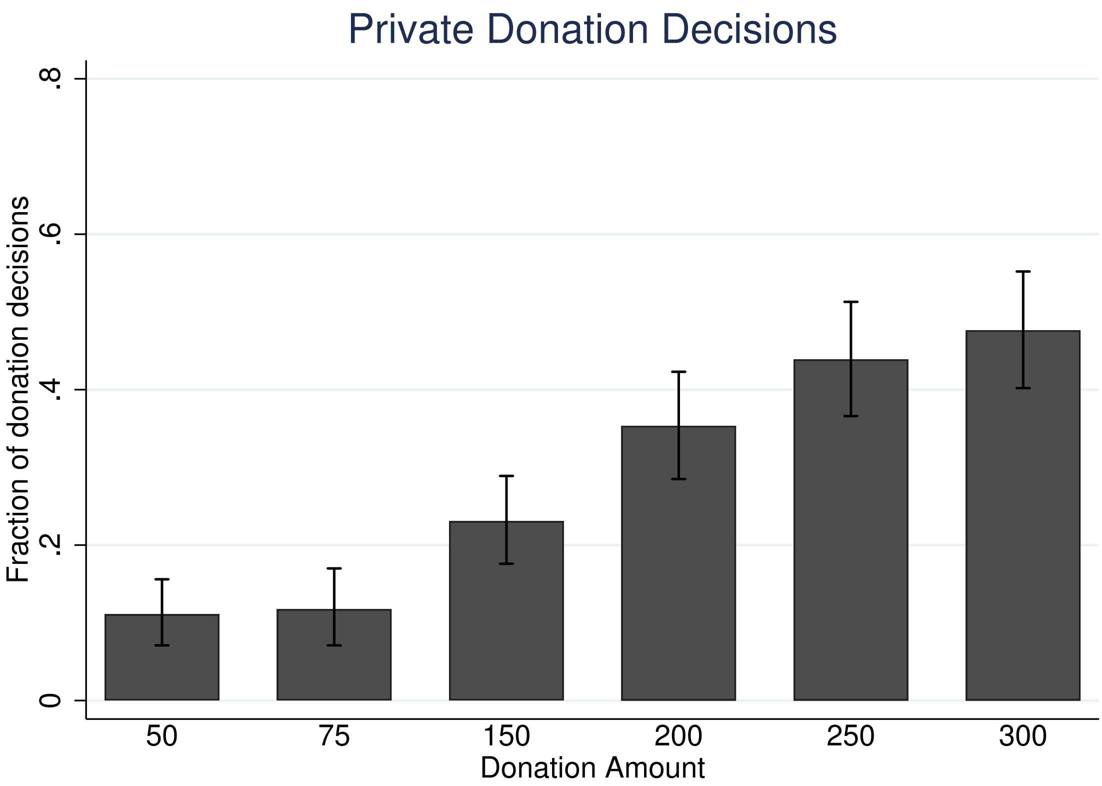


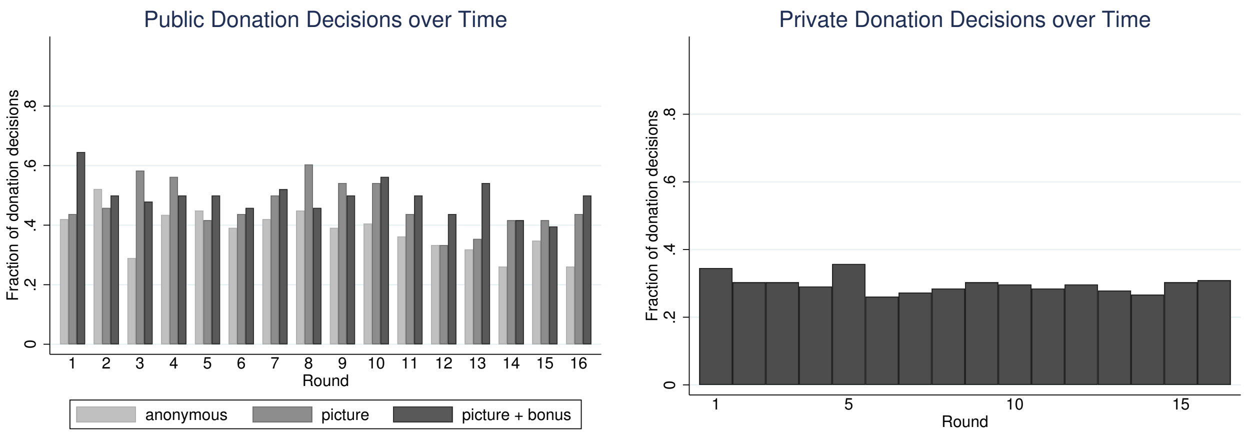
Figure A1: Private donation decisions.

Notes: Whiskers represent the 95% confidence intervals. Standard-errors are bootstrap with 10,000 repetitions and clustering at subject level.

Table A1: Private donation decisions

	(1)	(2)	(3)	(4)	(5)
public donation decision	0.171***	0.181***			
	(0.032)	(0.032)			
Efficiency Factor (X)		0.164***		0.160***	0.160***
in private situation		(0.017)		(0.017)	(0.017)
public donation unlikely ($p = 0.9$)			0.020		0.025
			(0.013)		(0.024)
Treatment: Anonymous				-0.034	-0.037
				(0.047)	(0.049)
Treatment: Picture+Bonus				0.074	0.079
				(0.059)	(0.059)
Treatment: Anonymous					0.005
× *public donation unlikely* ($p = 0.9$)					(0.031)
Treatment: Picture+Bonus					-0.010
× *public donation unlikely* ($p = 0.9$)					(0.031)
cons.	0.219***	-0.069**	0.282***	0.009	-0.004
	(0.020)	(0.023)	(0.022)	(0.037)	(0.038)
Obs.	2832	2832	2832	2832	2832
Clusters	165 subj.	165 subj.	165 subj.	165 subj.	165 subj.
R^2	0.035	0.139	0.000	0.109	0.110

Notes: OLS regressions on donation decisions in private situation. *public donation decision* is a dummy variable with value 1 in case the subjects decided in favor of donation in the public situation of the same round. Std. errors clustered at the subject level and depicted in parentheses. ***(**/*) significant at the 0.1 (1/5) percent level.



(a) Public donations, all treatments

(b) Private donations, all treatments

Figure A2: Share of donation decisions in different periods.

Table A2: Minimum detectable effect sizes

	Hypothesis	MDE (no clustering)	MDE (with clustering)	measured effect size
H1a	# donations ($p = 0.9$) > # donations ($p = 0.1$)	.046	.079	$.448 - .404 = .044$
	# donations (X_{high}) > # donations (X_{low})	.046	.088	$.582 - .257 = .325$
	# donations (Pic) > # donations ($Anonym$)	.054	.124	$.451 - .363 = .088$
	# donations ($Pic + Bonus$) > # donations (Pic)	.062	.126	$.495 - .451 = .043$
H2a	$E[\mu_i \mid a_i = 0, p = .1] > E[\mu_i \mid a_i = 0, p = .9]$	2.805	4.172	$32.413 - 32.747 = -.334$
	$E[\mu_i \mid a_i = 1, p = .1] > E[\mu_i \mid a_i = 1, p = .9]$	2.937	3.795	$46.634 - 45.663 = .971$
	$E[\mu_i \mid a_i = 0, Anonym] > E[\mu_i \mid a_i = 0, Pic]$	3.192	10.309	$28.393 - 35.405 = -7.012$
	$E[\mu_i \mid a_i = 1, Anonym] > E[\mu_i \mid a_i = 1, Pic]$	3.346	7.541	$45.403 - 43.759 = 1.643$
	$E[\mu_i \mid a_i = 0, Pic] > E[\mu_i \mid a_i = 0, Pic + Bonus]$	3.801	6.671	$35.405 - 35.944 = -.539$
	$E[\mu_i \mid a_i = 1, Pic] > E[\mu_i \mid a_i = 1, Pic + Bonus]$	3.838	8.065	$43.759 - 50.003 = -6.243$

Notes: Hypotheses 1a predict the share of donation decisions ($\in (0, 1)$). Hypotheses 2a predict the believed probability of a donation in the private decision in percent ($\in (0, 100)$). The minimum detectable effect sizes (MDE) are calculated using the standard deviation of the outcome variable in the data. Column MDE (no clustering) reports the minimum detectable effect sizes for a one-sided t-test given the number of observations ignoring the cluster dependencies in the data. Column MDE (with clustering) reports the minimum detectable effect size for a one-sided t-test given the number of observations taking clustering and the intra-cluster correlation into account. For hypothesis 1a the cluster levels are individual participants, for hypothesis 2a the cluster levels are matching groups.

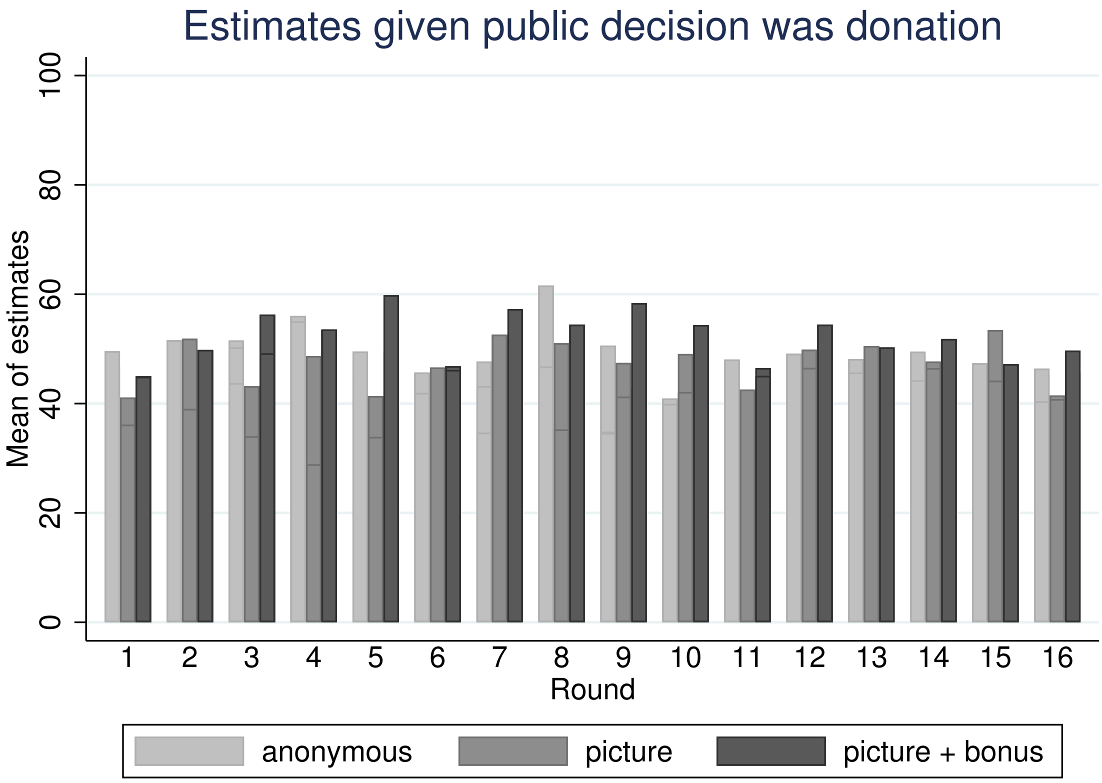


Figure A3: Average of observers' estimates given a public donation decision.

Notes: To ease comparison of the two sessions with 20 instead of 16 rounds, the rounds are depicted as the first 16 observed rounds (either corresponding to round 21 to 37 or to rounds 17 to 34 in the experiment).

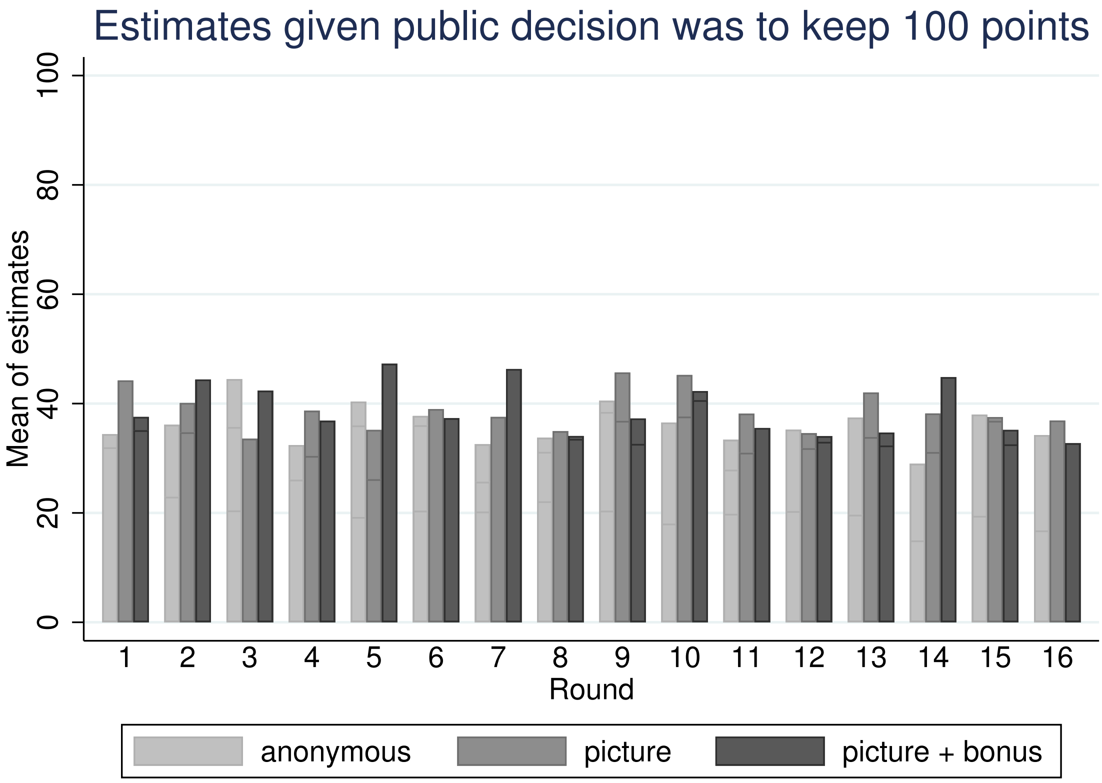


Figure A4: Average of observers' estimates given a public no-donation decision.

Notes: To ease comparison of the two sessions with 20 instead of 16 rounds, the rounds are depicted as the first 16 observed rounds (either corresponding to round 21 to 37 or to rounds 17 to 34 in the experiment).

B Decision Screens

Translations of the German texts (from top to bottom of each screen) are provided in the figure notes.

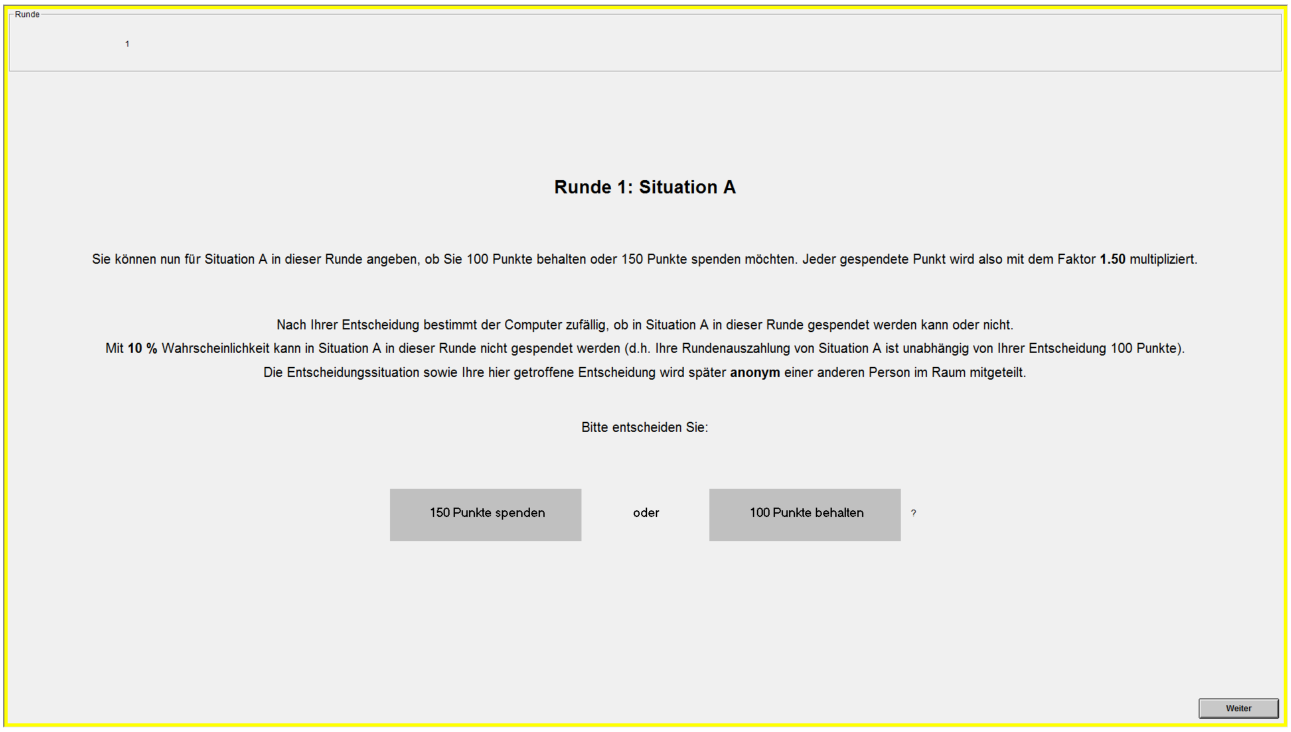


Figure B1: Treatment T1, Situation A.

Notes: "Round 1: Situation A. You can now state for Situation A in this round, whether you want to keep 100 points or donate 150 points. Every donated point will hence be multiplied by the factor 1.5. After your decision, the computer randomly determines whether donation can take place in Situation A of this round or not. With 10 % probability donation cannot take place in this round (i.e. your round payoff is 100 points, independently of your decision). The decision scenario as well as your decision will later be transmitted anonymously to another person in the room. Please decide: donate 150 points or keep 100 points."

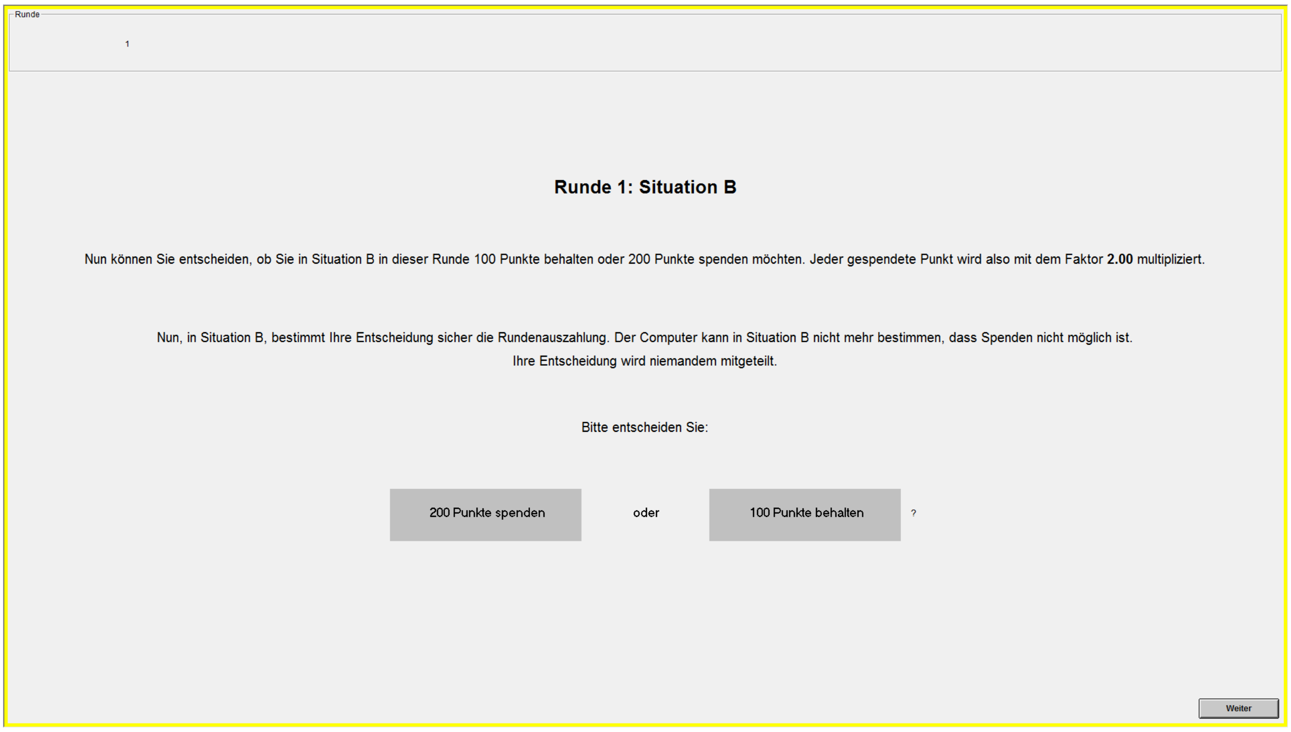


Figure B2: All Treatments, Situation B.

Notes: "Round 1: Situation B. You can now decide between keeping 100 points or donating 200 points. Hence, every donated point is multiplied by the factor 2.00. Now, in Situation B, your decision certainly determines the round payoff. In Situation B the computer cannot decide that donation is impossible. Your decision is not communicated to anyone. Please decide: Donating 200 points or keeping 100 points?"

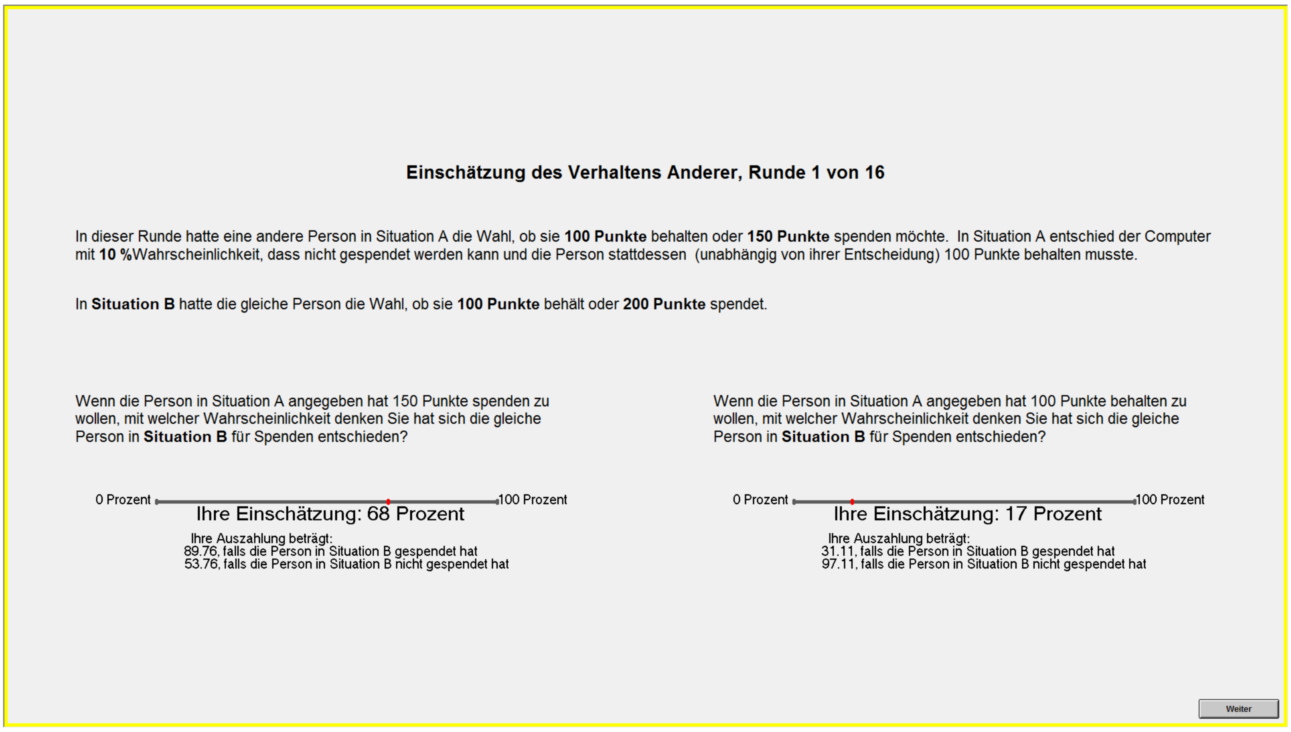


Figure B3: All treatments, Observer Screen.

Notes: "Estimating the behavior of others. In this round, another person could choose between keeping 100 points or donating 150 points. In Situation A, the computer decided with 10% probability that donation is not possible and that the person (independent of her/his decision) has to keep 100 points. In Situation B, the same person had the choice between keeping 100 points or donating 200 points. If the person indicated to donate 150 points in Situation A, what is the probability that the same person decided for donation in Situation B? Your estimation: 68%. Your payoff equals 89.76 in case the person donated in Situation B, your payoff equals 53.76 in case the person did not donate in Situation B. If the person indicated to keep 100 points in Situation A , what is the probability that the same person decided for donation in Situation B? Your estimation: 17%. Your payoff equals 31.11 in case the person donated in Situation B, your payoff equals 97.11 in case the person did not donate in Situation B."

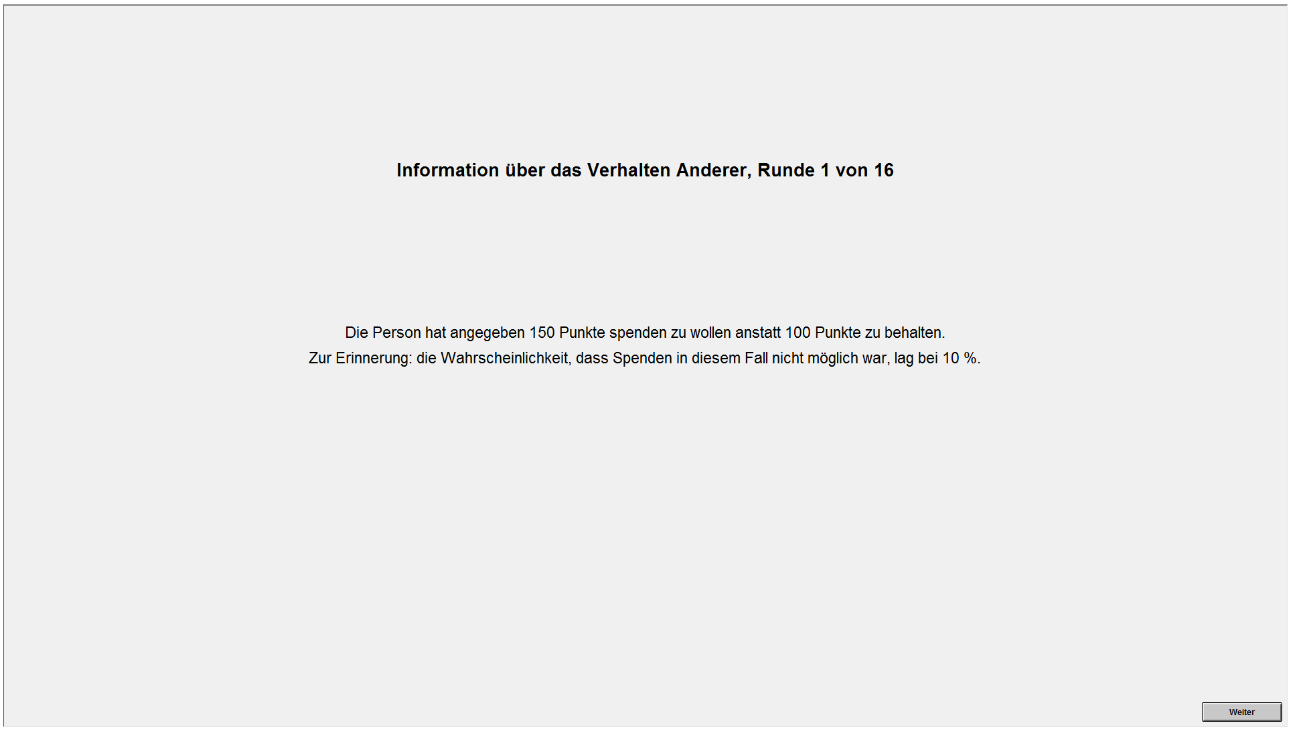


Figure B4: Treatment T1, Feedback of Observer.

Notes: "Information about the behavior of others. The person indicated the willingness to donate 150 points instead of keeping 100 points. Reminder: The probability that donation was not possible was in this case 10%."

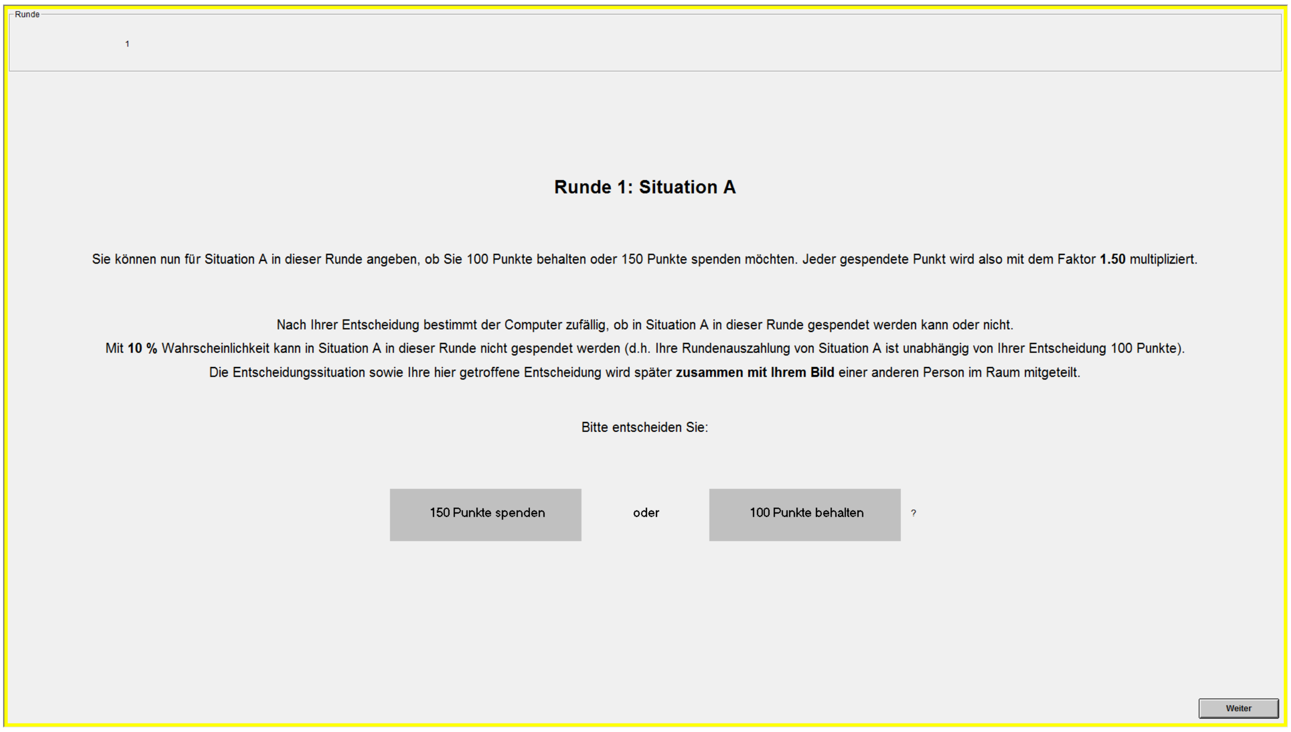


Figure B5: Treatment T2, Situation A.

Notes: "Round 1: Situation A. You can now state for Situation A of this round, whether you want to keep 100 points or donate 150 points. Every donated point will hence be multiplied by the factor 1.50. After your decision, the computer randomly determines whether donation can take place in Situation A of this round or not. With 10 % probability donation cannot take place in Situation A of this round (i.e. your round payoff in Situation A is 100 points independent of your decision). The decision scenario as well as your decision will later be transmitted together with your picture to another person in the room. Please decide: donate 150 points or keep 100 points."

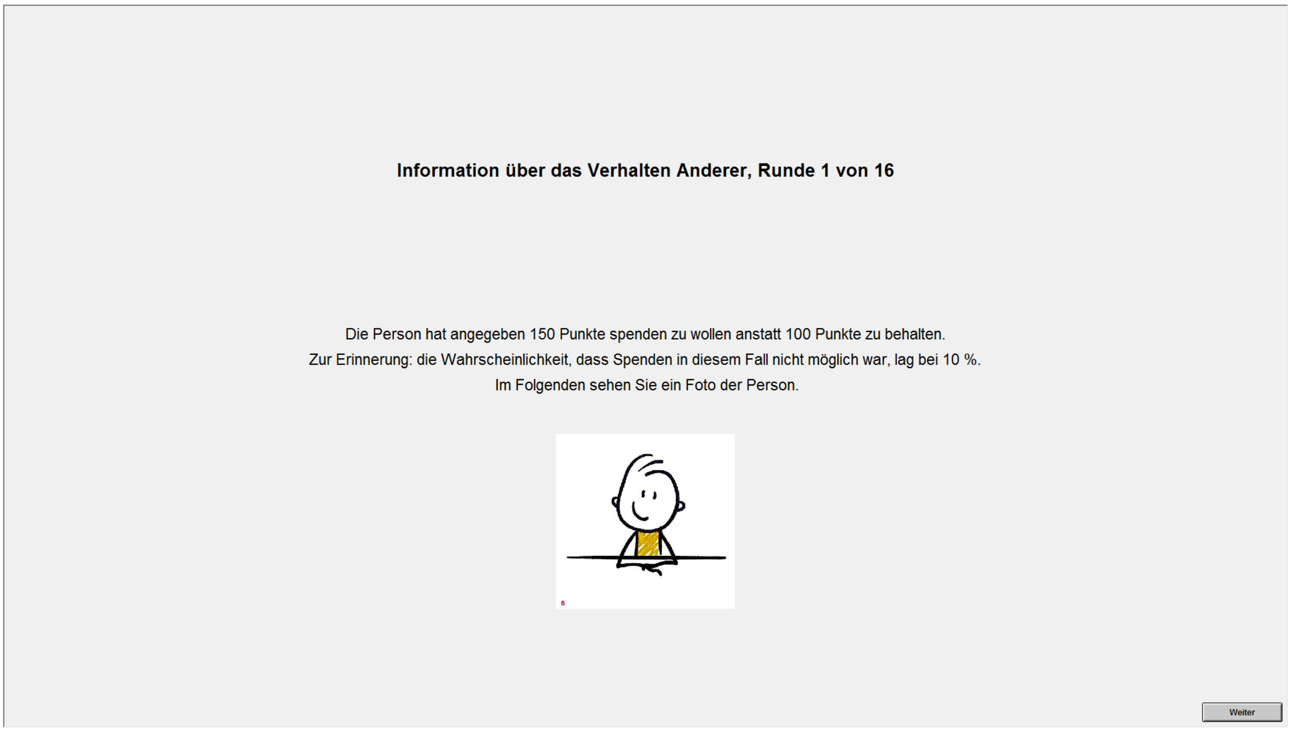


Figure B6: Treatment T2 and T3, Feedback of Observer.

Notes: "Information on the behavior of others, round 1 of 16. The person stated to be willing to donate 150 points instead of keeping 100 points. Reminder: the probability that donation was impossible was in this case 10 %. In the following, you can see a picture of this person."

Figure B7: Treatment T3, Situation A.

Notes: “Round 1: Situation A. You can now state for Situation A in this round, whether you want to keep 100 points or donate 150 points. Every donated point will hence be multiplied by the factor 1.50. After your decision, the computer randomly determines whether donation can take place in Situation A of this round or not. With 10 % probability donation cannot take place in Situation A of this round (i.e. your round payoff in Situation A is 100 points independent of your decision). The decision scenario as well as your decision will later be transmitted together with your picture to another person in the room. Please decide: donate 150 points or keep 100 points. On the basis of your decision above, another person in the room will later estimate the probability that you decided for donation in Situation B. The higher the estimate, the higher your round payoff.”

C Instructions

An English translation of the Instructions for the Anonymous treatment is provided below. The original German instructions of all three treatment conditions are part of the replication package and can be accessed via XXX.

Overview

Welcome to this Experiment. We ask you not to talk to the other participants during the course of the experiment and to switch off your cell phone and all other mobile devices.

For the participation in this session, you will be paid in cash at the end of the experiment and you can also arrange a donation. The payment will depend partly on your decisions, partly on the decisions of the other players and partly on chance. Your decisions and chance also determine the amount of your donation. For this reason, it is important that you understand the instructions before the experiment starts.

In this experiment, all interactions between the participants run through the computer in front of you. Your decisions are only saved together with your random ID number. Neither your name nor the names of other participants will be disclosed, neither today nor in future written evaluations.

In today's experiment you will act for several rounds in the role of a decision maker and several rounds in the role of an observer. At the end, individual rounds from each role are randomly selected as relevant for the payout. The decisions of the non-selected rounds will not be paid out. Your payment and your donation consist of the points that you earn and donate in the selected rounds, converted to euro, plus your show-up fee of 5 Euro, which you will receive in any case.

The conversion of the points in euro proceeds as follows: Every point is worth 1.5 cents, such that:

100 Points = 1,50 Euro

Every participant will be paid in private in such a way that the other participants cannot see how much you will have earned.

Experiment

The experiment consists of three parts. Part 1 (round 1-8) and part 2 (round 9 – 16) are identical in terms of their structure. In these first two parts, all participants act as **decision-makers**. In the third part of the experiment, all participants act as **observers**.

As a decision maker, you have the opportunity to make a donation to a non-profit organization in each round. At the end of the experiment, we will ask you to which of the three specified organizations (*Amnesty International*, *Ärzte ohne Grenzen* and *Deutsche Krebshilfe*) your donation should go. Information on the three organizations can be found on your table. No other participant can see which organization you choose.

Part 1 and 2: Decision-maker

In Part 1 (Round 1-8) and Part 2 (Round 9-16) you are the decision-maker. Each of the 16 rounds consists of **one decision in Situation A** and **one decision in Situation B**, i.e. you will make a total of 32 decisions. The following explanations refer to Part 1 of the experiment. You will receive the instructions for Part 2 (Round 9-16) of the experiment on your computer screen at the beginning of the 9th round.

The decision task: At the beginning of each of the 32 decisions, you will be given a budget of 100 points. You then have to decide whether to:

- **keep the 100 points** (and donate 0 points) or
- **donate 100*x points** (and keep 0 points).

Here we call x the factor. It determines the number of points that can be donated in the current decision task. **The factor (x) is determined anew and randomly for each person and each of the 32 decision tasks.** This means that it can vary over the rounds, but also between **Situation A** and **Situation B** of a round. The factor takes the values 0.5, 0.75, 1.5, 2, 2.5 or 3 with equal probability. Accordingly, either 50, 75, 150, 200, 250 or 300 points can be donated in the decision tasks and each of these values occurs with probability $\frac{1}{6}$ = 16,67%. Before each decision, you will be told the value of the factor.

The decision task described above is encountered by you in each round first in Situation A and then in Situation B. Situation A and Situation B differ in terms of (i) **the visibility of your decision** and (ii) the **impact** of your decision on your round payout.

<u>Situation A:</u> Decisions you make in **Situation A**,

- **will be shown to another person in the room (observer) in the third part of the experiment.** The observer is one person from a group which was randomly formed out of eight persons at the beginning of the experiment. He/she is randomly selected for each Situation A decision. This means: **not all (but possibly several) of your decisions in Situation A will be seen by the same person. Which of your Situation A decisions is seen by whom is random.** More details on the observer role can be found in Part 3.
- may be irrelevant to your round payoff: **With a probability of 90%, the computer determines that donating in Situation A is not possible in the current round. Then your round payout from Situation A (regardless of your decision) is 100 points.** You will only find out how the computer's random decision turned out at the end of the experiment.

Depending on your decision and the computer's random decision for the corresponding round, there are three possible outcomes in **Situation A**:

1. You decide to keep 100 points instead of donating 100*x points: In this case, the computer's random decision is irrelevant. Your round payout for **Situation A** is 100 points (and the donation amount for **Situation A** in this round is 0 points).
2. You decide to donate 100*x points and the computer determines that donating is not possible in this round (occurs with 90% probability): Your round payoff for **Situation A** is 100 points (and the donation amount for **Situation A** in this round is 0 points).
3. You decide to donate 100*x points and the computer decides that donating is possible in this round (occurs with 10% probability): Your donation amount for **Situation A** in this round is 100*x points (and your round payout for **Situation A** is 0 points).

<u>Situation B:</u> Decisions you make in **Situation B**,

- **will not be shown to other participants at any time.**
- **determine with certainty your round payout for Situation B**. That is, if you decide to keep 100 points in **Situation B**, your round payout for **Situation B** will be 100 points for sure (and your **Situation B** donation for this round will be 0 points). If you decide to donate 100*x points in **Situation B**, your donation to **Situation B** in this round will be 100*x points for sure (and your round payout from **Situation B** 0 points).

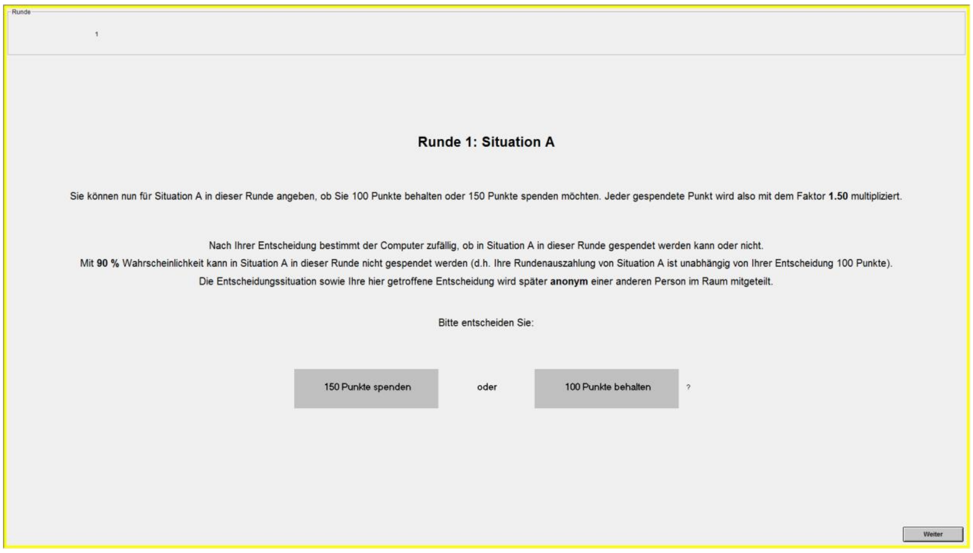


Figure 1: Example picture of a Situation A decision. The randomly determined factor in this example is 1.5.

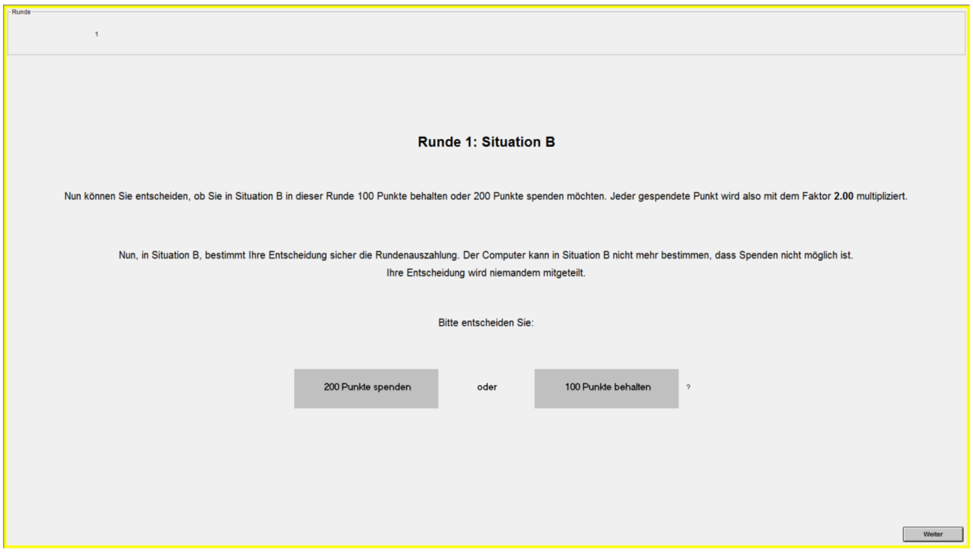


Figure 2: Example of a Situation B decision. The randomly determined factor in this example is 2.

Part 3: Observer

In the third part you are the observer. In each round, we describe the decision task of **Situation A** and the decision task of **Situation B** which a randomly selected person in the room faced in that round. This person is one of eight possible randomly selected persons (see description of **Situation A**).

We subsequently ask you to predict the behavior of the randomly selected person in **Situation B** based on the person's possible behavior in **Situation A**. In particular, it may be that the person

- decided in Situation A to **keep** 100 points (case 1)
- or decided in Situation A to **donate** 100*x points (case 2).

For both possible cases, we ask you to estimate the probability that the person decided to donate in Situation B. **You do not learn the person's actual behavior in Situation B. After you have entered your two estimates, we will only tell you how the person decided in Situation A.**

That is, in each round you make two estimates regarding another person's behavior in a specific **Situation B**: (i) one for the case that the person indicated that she/he wanted to donate in a specific **Situation A** and (ii) one for the case that the person indicated that she/he wanted to keep 100 points in a specific **Situation A**.

For your round payout only your estimate for the case that actually happened matters (i.e. the case for which the person in Situation A actually decided). The more accurately this estimate predicts the actual behavior in Situation B, the higher your round payout in part 3. Under the payout formula used in this experiment, you receive the highest expected payout as observer if you make your estimate truthfully. For example, if you think that the probability of a donation in **Situation B** is 60%, you maximize your expected payout if you also enter 60%.

The exact payout formula for the round payout as observer is:

- $100 - \frac{1}{100} * Prob(Donate\ in\ B)^2$ points if the other person did not donate in Situation B.
- $100 - \frac{1}{100} * \left(100 - Prob(Donate\ in\ B)\right)^2$ points if the other person donated in Situation B.

Where Prob(*Donate in B*) denotes the entered estimated probability of donation. The payout amount is always between 0 and 100 points and is displayed on the screen for the respective estimate.

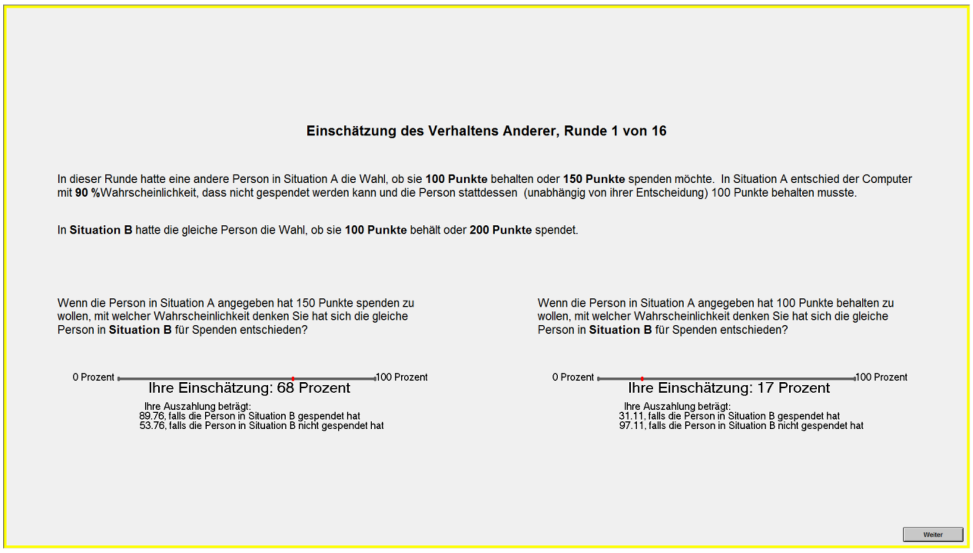


Figure 3: Exemplary picture of an assessment by the observer. In the observed round, the randomly determined factor was 1.5 in Situation A and 2 in Situation B.

After you have made your two estimates, you will be told the person's actual decision in **Situation A**; but not the person’s decision in **Situation B**.

Figure 4: Exemplary picture of the announcement of a Situation A- decision to the observer.

Determination of your payoff:

At the end of the experiment, your payout and the amount of your donation are determined as follows:

- **Situation A:** Two round payouts/donations from **Situation A** are randomly determined as relevant for your payoff: one from part 1 (round 1-8) and one from part 2 (round 9-16).
- **Situation B:** Two round payouts/donations from **Situation B** are randomly determined as relevant for your payoff: one from part 1 (round 1-8) and one from part 2 (round 9-16).
- **Part 3:** Two round payouts of the observer role are randomly determined as relevant for your payoff. To ensure that no inferences can be made about a decision in **Situation B**, it is not communicated which estimates were randomly selected.
- 5 euros for showing up on time.

Just as all payoffs are paid out directly at the end of the experiment, all donations are of course actually made. If you are interested, you are welcome to write to Lakelab (lakelab@uni-konstanz.de) after the experiment to see a donation receipt.

Questions?

Take your time to carefully go over the instructions again. If you have a question, please raise your hand. An experimenter will come to your cubicle to answer your question in person. If you think you understand everything well, click the ‘Start’ button on the screen. Next, you will be asked some control questions. The control questions aim to ensure that each participant has understood the instructions. The answers you give in these control questions do not affect your payoff. After answering the control questions, you can see and compare your answers to the right solution by clicking on “Next”. When all participants have answered the control questions and all remaining questions have been clarified, the experiment will begin.

Short description of the non-profit organization taken from their websites:

1. Amnesty International: „Amnesty International is the world's largest movement for human rights. Amnesty is independent of governments, parties, ideologies, economic interests and religions. In order to ensure this independence, we finance our human rights work solely from donations and membership fees. Our campaigns and actions are based on the principles of the Universal Declaration of Human Rights."

[...] „Amnesty was awarded the Nobel Peace Prize in 1977 for this work. The explanatory statement stated that Amnesty had a clear stance: "No to violence, torture and terrorism. On the other hand, a yes to the defence of human dignity and human rights". Amnesty advocates these values to this day."

2. Ärzte ohne Grenzen: „Ärzte ohne Grenzen provides emergency medical aid when the lives of many people are threatened in war zones or after natural disasters. We do not ask about origin, religion or political convictions, but only decide what our patient urgently needs.

- Your donation enables our worldwide emergency aid: Thanks to you, we can be where the world needs us - even in war and crisis areas.
- Ärzte ohne Grenzen is therefore 96 percent financed by donations from private individuals. In many politically charged contexts, we consciously reject government funds, because this is the only way we can secure our independence and impartiality. "

"In 1999, Ärzte ohne Grenzen received the Nobel Peace Prize in recognition of their humanitarian work."

3. Deutsche Krebshilfe: „Deutsche Krebshilfe was founded on September 25, 1974 by Dr. Mildred Scheel. The goal of the non-profit organization is to fight cancer in all its manifestations. According to the motto „Help. Research. Inform." the organization promotes projects to improve prevention, early detection, diagnosis, therapy, medical aftercare and psychosocial care, including cancer self-help. Deutsche Krebshilfe is the most important private donor in the field of cancer research in Germany."

"All activities of Deutsche Krebshilfe are only possible with the help of the population, our donors, but also the numerous people - doctors, scientists, personalities from all areas of society - who volunteer in our committees. Together we always focus on the concerns of people affected by cancer. "

Websites:

1. *https://**www.amnesty.de**/amnesty/wer-wir-sind*
2. *https://**www.aerzte-ohne-grenzen.de**/spende*
   *https://**www.aerzte-ohne-grenzen.de**/retten-sie-leben-spenden-sie-jetzt-fuer-aerzte-ohne-grenzen*
3. *https://**www.krebshilfe.de**/informieren/ueber-uns/deutsche-krebshilfe*
   *https://**www.krebshilfe.de**/spenden-aktiv-werden/spenden-service/ihre-spende-rettet-leben/was-passiert-mit-meiner-spende/*