# Cheap Signaling of Altruism

Moritz Janas
Department of Economics
University of Konstanz
Box 135
78457 Konstanz
Germany
moritz.janas@uni-konstanz.de

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

November 19, 2020
click here for the latest version.

**Abstract**

Signaling altruism is cheap, if it is unlikely that the intended, pro-social action will be effectuated. We adapt a model to predict how individuals decide if it is ex-ante unclear whether acting pro-socially is possible at all. The model consists of i) a preference for a prosocial reputation, ii) an unobservable valuation of helping others (altruism), iii) the effectiveness of acting pro-socially, and iv) how cheap signaling altruism is (i.e. how likely the intended prosocial act can be carried out). We solve the model for its perfect Bayesian equilibria and test the key predictions in a laboratory experiment. In the experiment, subjects face the choice of either receiving money themselves or donating it to a charity. Observers' perceptions, a key ingredient of the perfect Bayesian equilibrium, are elicited in an incentive compatible way. We vary how effective donation is and how likely it can take place across choices. Between-subjects we manipulate the strength of the incentive to appear pro-socially and whether it has an additional monetary component. The data confirms the predicted effect of cheaper signals, but only if the image benefit is not linked to monetary rewards. Contrary to theory, the actual social image - the beliefs of observers about the level of altruism of an individual - seems to be systematically biased, s.t. a selfish decision is not perceived more selfish, if the decision is unlikely to be carried out.

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

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

# 1 Introduction

Signaling games describe a situation in which costly actions (signals) can be used to transmit private information to others. In this paper individuals can transmit information about their private level of altruism by using *cheap* signals. We denote cheap signals as signals that are costless with exogenous, positive probability. In our scenario the cheap signal is the decision to engage in a costly pro-social act which may, however, turn out to be impossible and thus costless ex post. Pro-social decisions that may turn out to be costless after the decision (i.e. after the transmission of the signal) are widely observable in the real world. One example are conditional favors. Suppose e.g. that an acquaintance asks you whether he could stay a few nights at your place in case he gets accepted at a specific conference. You both know, however, that he only gets accepted with a particular chance. So, even if you assure your hospitality, you might not have to host your acquaintance and thus you might not have any costs. An example of a cheap signaling game with more severe consequences is stem cell donation. The willingness to be tested and added to the database signals altruism to some degree, the probability to really help someone (and to incur the costs of helping someone), however, depends on the exogenous probability that a match is found.

We study, theoretically and experimentally, how individuals behave when signaling altruism is cheap. Specifically, we investigate how the probability that pro-social decisions are infeasible influences how individuals decide and how observers evaluate these decisions. We also explore which role different types of reputation play in this context. That is, whether people react more or less strongly to cheapened signaling costs if the reputational incentive is high or low.

Although people frequently encounter situations in which signaling altruism is cheap, there exist only few papers which study such decision contexts (e.g. Grossman, 2015; Falk and Tirole, 2016; Carter and Guerette, 1992; Tyran, 2004).[1] All of them 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 reputational incentives. To test this conjecture, we systematically vary the type of reputational incentive in our experiment. Moreover, we elicit how decisions are perceived by others 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 decisions are actually perceived by others. Our paper aims to contribute with respect to perceptions and the type of reputational incentive, but

[1] We refer here to the subsection "The cheap signaling effect" in the working paper version of Falk and Tirole (2016).

also tries to deepen the understanding of situations in which signaling altruism is cheap in general.

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

We first use a model based on Bénabou and Tirole (2006), Bénabou et al. (2018a) and Grossman (2015) to replicate the core prediction that more people decide pro-socially when signaling altruism becomes cheaper. From this model we also derive comparative statics predictions for how perceptions should change when the signaling costs decrease and for how the change in decisions and perceptions should depend on the reputational incentive. We then test these predictions in a laboratory experiment. In the experiment subjects can decide between keeping money or donating to a charity instead. This choice is made observable to other subjects. However, sometimes donation turns out to be impossible ex post, i.e. after the decision-maker made his choice. We use a (3x2)-design in which we manipulate the type of reputational incentive between subjects (choices are made observable in an anonymous manner vs. with picture of the decision-maker vs. with picture and there is a monetary bonus for a pro-social appearance) and the signaling costs within subjects (slightly cheap vs. very cheap). By comparing decisions and perceptions in settings with different signaling costs, we can draw inferences on question one and three. The comparison of decisions and perceptions across different reputational incentives gives us insights on the second question.

Our experimental findings can be summarized as follows: As postulated by our model we find that more individuals decide pro-socially if signaling altruism is cheaper. However, this is only true if the reputational incentive is not too strong. In the treatment in which a pro-social appearance earned decision-makers a monetary bonus, the fraction of pro-social decisions does not increase in response to cheapened signaling costs. Indeed in this treatment cheapened signaling costs influence decisions significantly less. This observation is only partly in line with our model. While our model does predict a dependence on reputational incentives, it forecasts that strengthening reputational incentives would induce a larger (and not a smaller) increase in pro-social decisions. Moreover, it predicts that observers account for the cheapness of signaling by taking pro-social signals less and selfish signals more at face value. Our experimental data offers though finds a systematic bias therein: While cheapened signaling costs decrease the perceptions of observers after a donation decision, they do not decrease the perception of observers after a non-donation decision.

## 2 Related Literature

Most closely related are papers which study pro-social (or fair) decisions in a context in which not only image, but also pivotality plays a role (e.g. Carter and Guerette, 1992; Tyran, 2004; Grossman, 2015; Exley, 2016; Falk and Szech, 2017; Andreoni and Bernheim, 2009; Bénabou et al., 2018a). However, only in Carter and Guerette (1992), Tyran (2004), Grossman (2015), Falk and Tirole (2016) and Bénabou et al. (2018a) the uncertainty that a costly pro-social decision is implemented cheapens signaling altruism. The uncertainty in Exley (2016), Falk and Szech (2017) and Andreoni and Bernheim (2009) actually creates an excuse to decide selfishly making a pro-social decision not more, but less attractive. 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., 2018b). Overall, we are not aware of any experimental paper that specifically tests the perception of people's pro-social decisions when a) the signaling costs change, and/or b) the reputational effect differs (weak vs. strong social reputation). Papers in which signaling costs are cheapened by reduced pivotality (Carter and Guerette, 1992; Fischer, 1996; Tyran, 2004; Grossman, 2015) focus on how the inclination to decide (vote) pro-socially changes when the pivotality of a pro-social decision (vote) is decreased. Evidence regarding this question is mixed.[2] Despite the theoretical importance of perceptions, these previous experiments did not explore how perceptions are affected by a decline of pivotality. We fill this gap. Moreover, by systematically varying the strength of social image concerns, we explore which role reputational incentives play in the context of reduced pivotality and offer a potential explanation for why previous evidence is mixed. In our theoretical part we build on existing models (mainly on Bénabou and Tirole, 2006; Falk and Tirole, 2016; Grossman, 2015), but tailor it precisely to our research questions.

## 3 Theory

### 3.1 The model

We use the signaling model of Bénabou and Tirole (2006) and, similarly as in Grossman (2015) and in Falk and Tirole (2016), extend it by an exogenous probability that the pro-social act does not take place. More specifically, individuals first publicly decide whether to engage in a costly pro-social act ($a_i = 1$) or not ($a_i = 0$) and then - after the decision is made - an exogenous factor determines whether the pro-social act is possible at all.

[2] Carter and Guerette (1992) find weakly significant evidence for more donation votes in case 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 prosocial choices for lower pivotality among these subjects. Tyran (2004) finds evidence for a bandwagon effect, i.e. subjects behaving as they expect other's to behave, which contradicts the idea of lower approval for a prosocial decision in case the pivotality within a committee increases. Grossman (2015) finds no significant influence of lower pivotality on donation behavior, despite a rather large sample. For a detailed literature review on expressive voting, see Tyran et al. (2016).

With probability $p$ pro-social acts cannot be implemented such that the individual can not behave pro-socially even if she decided for it. The probability $0 \leq p \leq 1$ is commonly known to all individuals. Consequently, not only the individual who decides upon the pro-social matter, but also the individuals who observe this decision know $p$.

Individuals differ in their degree of altruism $\mu_i \sim U(0,1)$ which is private information. Only the distribution of altruism is public information. The degree of altruism crucially 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 $\mu_i$ receives utility $\mu_i b \geq 0$ from creating a benefit $b$ to others. This utility sets off part of the direct costs ($c > 0$) which are associated with the pro-social act. The net costs of an (implemented) pro-social act are therefore $c - b\mu_i \geq 0$ and lower for individuals with a higher degree of altruism.[3]

Moreover, individuals are image-concerned in the sense that appearing altruistic to others generates positive utility.[4] Others update their beliefs about an individual's level of altruism based on the individual's decision ($a_i$) and the decision setting ($p, c, b$). The belief after observing decision $a_i \in \{0,1\}$ is given by $E[\mu_i|a_i, p, b, c, \lambda]$. The strength of image concerns is captured in the parameter $\lambda > 0$ which is for simplicity identical for all individuals.

Equation 1 summarizes the expected utility of an individual:

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

This creates a signaling framework, as there is a tradeoff between the costs (net costs) and the benefits (altruistic image) of deciding pro-socially. In the following, we first apply the model to the specific decision-setting of the experiment. Then we solve the game for its equilibria and elaborate how making the implementation of pro-social decisions unlikely affects the decisions and the perceptions of these in equilibrium.

## 3.2 The decision situation in the experiment

In the experiment the costly pro-social act, in which people can engage or not, is a donation to charity. Specifically, participants have to decide between either donating $100 \cdot X$ points or keeping 100 points. Thereby, $X$ can be one of six possible values: $X \in \{0.5; 0.75; 1.5; 2; 2.5; 3\}$. We refer to $X$ as efficiency factor, because it shows how

[3]In principle also $(c - b\mu_i) \leq 0$ may hold, i.e. donation is not costly. The interesting case is, however, when donation is costly, because then decisions may change when $p$ changes. Otherwise one should always donate independent of $p$.

[4]This utility for being seen as pro-social can source from a valuation for an altruistic image per se and/or future benefits linked to such an image. For our purpose, the exact source does not matter (as in Andreoni and Bernheim, 2009).

efficient giving up 100 points is. The individual knows the exact value of $X$. That is, she knows exactly how much would be donated in case she foregoes 100 points and donation is implemented. Moreover, she knows that with probability $p$ donation cannot take place and she must keep 100 points - independent of how she decided. Therefore it is clear that a donation causes direct costs of $c = 100$ points and a benefit of $b = 100 \cdot X$ points for charity. Moreover, donations can be realized with probability $(1-p)$ only, which is also known. After the individual decided, her decision and the exact decision setting $(p, X)$ is revealed to another person. This is common knowledge. Adopting all other assumptions from above, the expected utility becomes:

$$EU(\mu_i, a_i, p, X, \lambda) = \lambda E[\mu_i | a_i, p, X, \lambda] - a_i(1-p)(100 - 100X\mu_i)$$

### 3.3 Equilibria

We use the concept of Perfect Bayesian Equilibrium (PBE) to calculate all threshold separating equilibria in which all types above a particular threshold decide to donate. In a PBE each individual behaves optimally given her beliefs and these beliefs are consistent with the strategies that are chosen in equilibrium.

When deciding whether to donate $100X$ or keep 100 points, the individual trades off the relative image benefit against the relative expected costs. The relative image benefit is shaped by the strategies of all individuals. Given that individuals with $\mu_i \geq \bar{\mu}$ decide in favor of donation and individuals with $\mu_i < \bar{\mu}$ decide against donation, the threshold type, $\bar{\mu}$ (for whom the net image gain just equals the net costs of donation) is implicitly defined by:

$$\lambda \left( E[\mu_i | \mu_i \geq \bar{\mu}] - E[\mu_i | \mu_i < \bar{\mu}] \right) = (1-p) \cdot (1 - X\bar{\mu})\, 100$$

$$\Leftrightarrow \lambda \frac{\bar{\mu}+1}{2} - \lambda \frac{0+\bar{\mu}}{2} = (1-p) \cdot (1 - X\bar{\mu})\, 100$$

$$\Leftrightarrow \lambda \frac{1}{2} = (1-p) \cdot (1 - X\bar{\mu})\, 100$$

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

Note that more people decide for donation if $\bar{\mu}$ is lower, as individuals with $\mu_i > \bar{\mu}$ decide for donation.

In the following, we first provide predictions regarding our research questions (Proposition 1-3) and then predictions for other aspects of the setting (summarized in Proposition 4). The derivations and proofs leading to Proposition 4 can be found in the Appendix.

**Proposition 1:** *More individuals decide for donation if donation is less likely possible, i.e. p increases. In other words, there are more pro-social decisions if signaling altruism is cheaper.*

**Proof.**

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

Proposition 1 can be understood by realizing that the expected costs of donation decisions decrease if donation is less likely possible ($p$ increases). If $p$ is high, it is rather likely that one can decide for donation but still keep the money. Thus, committing to donation is often costless. Yet, it still generates the same relative image benefit.[5] Therefore deciding for donation becomes more attractive. Now also less altruistic individuals, for whom it was (due to their relatively low altruism degree) initially too costly to decide for donation, switch to donation decisions. The number of donation decisions increases.

Note that the strength of image concerns $\lambda$ defines how much individuals value the relative image benefit of switching to a donation decision, $\lambda\left(E[\mu_i|\mu_i \geq \bar{\mu}] - E[\mu_i|\mu_i < \bar{\mu}]\right)$. If the reputational incentives are high ($\lambda$ high) image gains are more valuable. We therefore expect that individuals exploit the chance to cheaply improve their image more, the more they appreciate image gains. Consequently, we expect that cheapened signaling costs (due to increases in $p$) cause more switches to donation decisions when $\lambda$ is higher. In Proposition 2 we put forward this expectation.

**Proposition 2**: *Pro-social decisions increase more strongly when p increases if the reputational incentive is higher, i.e. $\lambda$ is higher.*

**Proof.**

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

Last, changes in the signaling costs should not only affect decisions but also how these decisions are perceived by others. Bayesian updaters should understand how cheapened signaling costs affect the incentive to decide for donation and adapt their beliefs accord-

[5] Since we assume uniformly distributed types, the relative image benefit is constant in our model. For other distribution assumptions the relative image benefit may however also decrease when $p$ increases. Yet, 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)) Proposition 1 holds.

ingly. This consideration motivates Proposition 3.

**Proposition 3:** *The perceived degree of altruism following a donation decision ($a_i = 1$)* ***and*** *following a keep decision ($a_i = 0$) is higher, the more likely donation is possible, i.e. the lower $p$.*

**Proof.** The equilibrium perceptions of donation and keep decisions are $\frac{\bar{\mu}+1}{2}$ and $\frac{0+\bar{\mu}}{2}$, respectively. Note that as $p$ increases, $\bar{\mu}$ decreases. This in turn implies that $\frac{\bar{\mu}+1}{2}$ and $\frac{0+\bar{\mu}}{2}$ decrease.

The intuition behind Proposition 3 is that if it is very cheap to signal altruism ($p$ is high), observers 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 person did not decide pro-socially, even though it would have been very cheap to do so. In Figure 1 we illustrate how the perceptions of decisions, $E[\mu_i|a_i = 1]$, change when the signaling cost go from slightly cheap to very cheap. That is, we visualize the insights of Proposition 3. However, also the insights of Proposition 1 can be found in the figure: when moving from slightly cheap to very cheap signaling the threshold type decreases such that more individuals decide for donation.

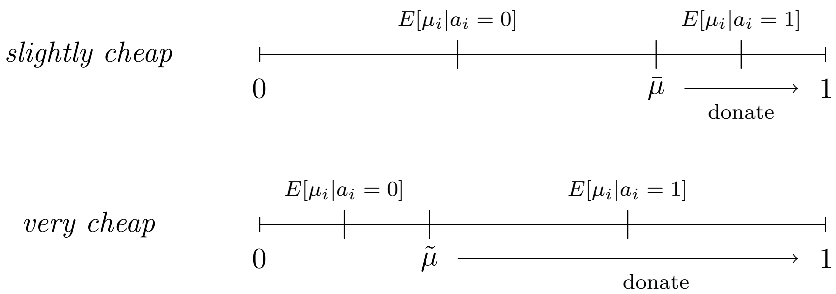



Figure 1: Illustration of threshold types ($\bar{\mu}, \tilde{\mu}$) and beliefs about an individual's altruism type ($E[\mu_i|a_i = 0]$, $E[\mu_i|a_i = 1]$) when signaling is slightly cheap ($p$ low) vs. very cheap ($p$ high).

Our model also generates other, rather intuitive predictions which we summarize in Proposition 4. These predictions are not the main focus of this paper. However, because they offer further guidance on what behavior one could (reasonably) expect in our decision setting, we derive them in Proposition 4 and will also test them in the experiment. The short proofs leading to Proposition 4 can be found in the Appendix.

**Proposition 4:**

**4.1** *The perceived altruism, $E[\mu_i|a_i, p, X, \lambda]$, is* ***(a)*** *lower if in the observed decision donation was more effective, i.e. $X$ is higher and* ***(b)*** *higher if a person decided for* *donation ($a_i = 1$) than if not ($a_i = 0$).*

**4.2** *More persons decide for donation **(a)** if donation is more effective, i.e. $X$ is higher and **(b)** if the reputational incentive, i.e. $\lambda$, is higher.*

# 4 Experiment

## 4.1 Experimental Design

**Overview.** Table 1 gives a schematic overview of our experimental design. We use a $2 \times 3$-design in which we manipulate the signaling costs within-subjects and the reputational incentive between-subjects. Signaling in *Situation A* can be either slighty cheap or very cheap and the reputational incentive in *Situation A* can be either weak, strong or strong plus monetary.

Table 1: Treatment overview. $p$ denotes the probability that donation is impossible.

| Treatments | slightly cheap: $p = 0.1$ | very cheap: $p = 0.9$ |
|---|---|---|
| T1: anonymous (weak) | T1-slightly | T1-very |
| T2: with picture (strong) | T2-slightly | T2-very |
| T3: picture + bonus (strong + monetary) | T3-slightly | T3-very |

**Details.** The experiment consists of two parts of equal length. In the first part (round $1-16$) subjects are decision-makers; in the second part (round $17-32$) they are observers.[6]

In the role of the decision-maker subjects are every round confronted with two decision problems: first with the main decision problem in which signaling altruism is cheap (called *Situation A*), and then with the supplementary one (called *Situation B*). We include *Situation B* decisions to test how *Situation A*-decisions are perceived. In both decision problems subjects have to decide whether to take 100 points or do a donation instead. The donation amount is determined by the efficiency factor $X = \{0.5, 0.75, 1.5, 2.0, 2.5, 3.0\}$ and can be either $50, 75, 150, 200, 250$ or $300$ points. The efficiency factor is chosen randomly for each of the 32 decision-problems.[7] For *Situation B* subjects decide in private whether to donate or not and these decisions are always implemented. However, in *Situation A* donations are not always possible: with probability $p$ a donation decision cannot be implemented. In this case, subjects must keep the 100 points irrespective of their actual decision. That is, if a subject decides in favor for donation in *Situation A* this results only with probability $(1 - p)$ in a donation of $100 \cdot X$ points and the costs associated with it (i.e. forgoing 100 points). We explained this in detail to subjects and they also knew $p$ and thus the exact

[6] In the pilot and session 1 each part lasted 20 rounds. Since 20 rounds turned out to be too time-consuming, we cut it down to 16 rounds for the remaining sessions.

[7] Therefore, the donation amount could differ across rounds, but also between *Situation A* and *Situation B* of the same round.

probability with which their decisions were implemented. Moreover, decisions made in *Situation A* are in the second part of the experiment revealed to other subjects (*observers*). The way of revelation depends on the reputational treatment.

After playing 16 rounds as *decision-maker*, subjects act as *observers* and keep this role for the remaining 16 rounds. In this role they observe every round a randomly chosen subject (of their matching group[8]) and are then asked to make an estimate regarding the behavior of this subject. More specifically, they first observe the specific *Situation A* and *Situation B* which this person faced in the particular round, i.e. the donation amount $100 \cdot X$ of each situation and the probability $p$ that donation was impossible in *Situation A*. Then, they estimate the probability that this person decided for donation in *Situation B* (i) given that the person decided in favor of donation in *Situation A* and (ii) given that the person decided against donation in *Situation A*. Only after observers entered these two estimates, it is revealed how the observed person actually decided in *Situation A*. This procedure has two advantages. First, the strategy method allows us to gather more data on how decisions are perceived. Second, showing the actual decision only ex post, i.e. after the estimates are entered, rules out unintended correlations between the picture of the decision maker they see and the estimate in $T2$ and $T3$. It is important to stress that *observers* never see the actual *Situation B*-decisions and that this is absolutely clear to subjects. This allows us to use *Situation B*-decisions as proxy for subjects' actual altruism (excluding signaling motives) and the *observers'* estimates as proxy for how altruistic a person is perceived to be. We incentivized these estimates by using a quadratic scoring rule which ensured that honestly reporting one's belief about how the observed person behaved in *Situation B* maximized one's expected payoff.[9]

To test how cheapened signaling costs affect decisions and estimates, we manipulate the probability $p$ that donation is impossible in *Situation A* within-subjects. $p$ can be either 0.9 or 0.1. For $p = 0.9$ the costs of a donation decision hardly never realize which renders signaling altruism cheaper than if $p = 0.1$. Moreover, we manipulate the reputational incentive by varying the way in which *Situation A*-decisions are presented to observers. Overall, we have three different reputational treatments and each subject participates in only one of these (between-subject design). In the first treatment ($T1$) observers see *Stituation A*-decisions in an anonymous manner, i.e. observers only know that "one participant decided for...". In the second treatment ($T2$) observers see *Situation A*-decisions together with a picture of the decision-maker. To this end we use photographs that we took after participants were seated in the lab. The photographs were deleted directly after the experiment and never connected to the participants' name. In the third treatment ($T3$) we connect *Situation A*- decisions to pictures (like in $T2$) and additionally pay a monetary

[8]Matching groups were randomly formed in the beginning of each session and consisted of eight persons. Since we varied the reputational incentive across sessions, subjects of the same matching group encountered the same reputational treatment.

[9]We only paid the estimate that belonged to the *Situation A*-decision which was actually made.

bonus for a pro-social appearance vis-à-vis the observer. Thereby, we strengthen the reputational incentive further but also change its nature. While in $T1$ and $T2$ an altruistic appearance is only beneficial in affective terms, in $T3$ it can also be used to increase one's payoff.

In the end of the experiment subjects choose in private whether their donation(s) should go to Amnesty International, Medecins Sans Frontieres or German Cancer Aid.[10] We offered a list of charities to increase the chance that each subject finds at least one charity which he or she perceives deserving. After choosing their preferred charity, subjects received their payment in private and the donated amount was directly transferred to the indicated charity. The payment consisted of a flat payment (5 Euro), two randomly selected estimates made as *observer* and four decisions made as *decision-maker* (a randomly chosen *Situation A*- and *Situation B*-decision for round $1-8$ and round $9-16$ each).

## 4.2 Procedure

We programmed the experiment with ztree (Fischbacher, 2007) and used ORSEE (Greiner, 2015) to recruit subjects.[11] In total, we conducted 7 sessions with 165 participants in November and December 2018.[12] In sessions that included pictures, we informed about the picture before entering the laboratory and gave potential participants the chance to leave directly. The announcement clarified that (i) we will take a picture in the beginning (ii) this picture will be shown together with certain decisions to other participants (iii) it will always be clear which decisions will be shown to others and (iv) pictures were never connected to names and deleted directly after the experiment. In all treatments we explained in-depth how exactly *Situation A*-decisions were revealed.

To control for order effects we varied whether the probability that donations are impossible, $p$, changed from high to low or low to high. In some sessions $p$ was first 0.9 (in round $1-8$) and then 0.1 (in round $9-16$); in other sessions it was first 0.1 and then 0.9. Moreover, we gave subjects in the beginning only the instructions for round $1-8$ and presented the instructions for the remaining rounds $(9-16)$ directly on the screen. Consequently, subjects did not know in advance that there will be a change in probabilities.

## 4.3 Hypotheses

All of our hypotheses build on the propositions in section 3.3. We start with hypotheses regarding our research questions (1-3) and then offer supplementary hypotheses (4).

[10] We chose charities that rank high in credibility and respectability surveys.
[11] Screenshots of the decision screens can be found in Appendix.
[12] The average age was 22.18 ($sd = 2.59$) and 37.58% ($sd = 0.49$) of the participants were male.

**Hypothesis 1.** Donation decisions in *Situation A* are more frequent if the probability that donation is impossible is $p = 0.9$ than if it is $p = 0.1$.

**Hypothesis 2.** The effect of $p$ on decisions (as described above) is weakest in $T1$ and strongest in $T3$.

**Hypothesis 3.** An observer's estimate (i.e. the incentivized belief that a person decided for donation in *Situation B* given the decision in *Situation A*) is higher if the observed person faced $p = 0.1$ than if he/she faced $p = 0.9$. This holds independent of the decision in *Situation A*.

**Hypothesis 4.**

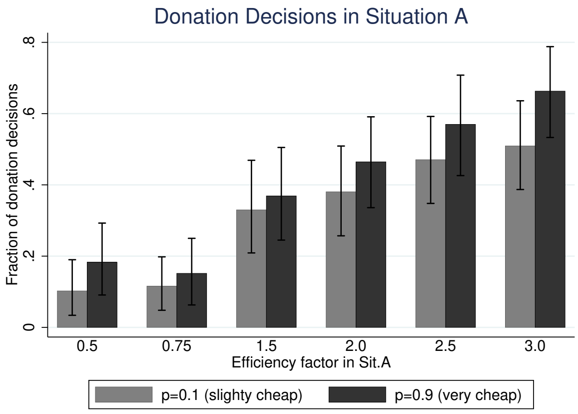
**4.1** Observers' estimates are lower **(a)** if the efficiency factor of *Situation A* is higher and **(b)** if the observed person kept the 100 points in *Situation A* (instead of deciding for donation).

**4.2** The fraction of donation decisions **(a)** in *Situation A and B* is higher if the respective efficiency factor is higher and **(b)** in *Situation A* is lowest in $T1$ and highest in $T3$.

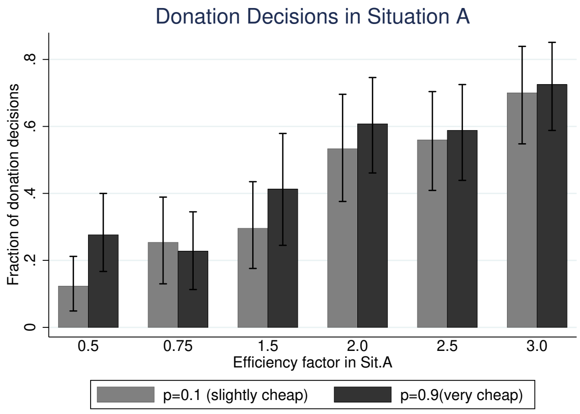
# 5 Results

## 5.1 The effect of cheapened signaling costs

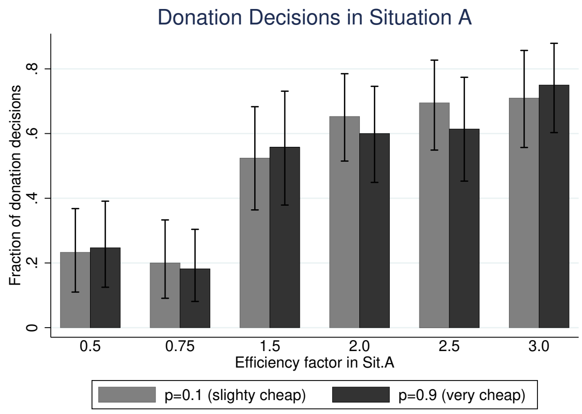
Hypotheses **1-3** make specific predictions about how subject's behavior (decisions and estimates) change when signaling altruism gets cheaper and how this change depends on the reputational treatment. We now present how subjects actually behaved. In hypothesis **1** we surmised that subjects will exploit the opportunity to cheaply signal altruism and decide more often for donation than they would otherwise do. That is, we expected that more subjects will publicly indicate their willingness to donate if the chance that they must take the 100 points *nevertheless*, i.e. the costly donation is infeasible, is very high ($p = 0.9$). In Figure 2 (a) one can see that participants, whose *Situation A*-decisions were disclosed in an anonymous manner, indeed reacted rather strongly to cheapened signaling costs. Independent of whether a donation decision led to $50, 75, 150, 200, 250$ or $300$ points if implemented, we always observed more donation decisions if deciding for donation was very cheap ($p = 0.9$) than if it was only slightly cheap ($p = 0.1$). On average, cheapening signaling costs increased the percentage of donation decisions by 8 percentage points ($p < 0.01$, Table 2). A similar pattern is also observable in $T2$, in which observers saw the *Situation A*-decisions together with a photo of the decision-maker, albeit somewhat weaker than in $T1$. In $T2$ cheapened signaling costs induced, on average, a 6 percentage points higher fraction of donation decisions. With respect to the effect of cheapened signaling costs only $T3$ builds an exception. In this treatment cheapening signaling costs did not significantly change decisions (Table 2). In Figure 2 (c) it is straightforward to see that behavior does not change in a systematic way if the probability that donation is impossible increases from 0.1 to 0.9. Seemingly, paying additionally a monetary bonus for a pro-social appearance prevents an effect to occur.




(a) T1: anonymous




(b) T2: with picture




(c) T3: with picture + bonus

Figure 2: This figure illustrates the frequency of donation decisions in Situation A depending on the efficiency factor in Sit. A (the potential donation amount is efficiency factor $\cdot 100$ points), the reputational treatment ($T1, T2, T3$) and signaling costs in Sit. A ($p = 0.1, p = 0.9$). Bars in light gray depict the share of donation decisions if signaling was slightly cheap, i.e. donation was impossible with probability $p = 0.1$. Bars in dark gray stand for rounds in which signaling was very cheap, i.e. $p = 0.9$. The depicted 95% confidence intervals are based on a bootstrap with 10,000 repetitions and clustering at subject level.

**Result 1:** *More individuals decided for donation if donation was less likely possible. However, only in the treatments in which decisions were revealed in an anonymous manner or with pictures.*

These results motivate the conjecture that the effect of cheapened signaling cost be-

Table 2: Decisions in Situation A by treatment

| | anonymous | | picture | | picture+bonus | |
|---|---|---|---|---|---|---|
| | OLS | Logit | OLS | Logit | OLS | Logit |
| Efficiency factor | 0.19*** | 0.91*** | 0.20*** | 0.91*** | 0.22*** | 0.96*** |
| in Sit.A | (0.03) | (0.13) | (0.03) | (0.16) | (0.04) | (0.19) |
| Donation in Sit.A | 0.08*** | 0.42*** | 0.06 | 0.30* | -0.01 | -0.06 |
| unlikely (p=0.9) | (0.03) | (0.15) | (0.04) | (0.18) | (0.04) | (0.19) |
| Clusters | 69 subj. | 69 subj. | 48 subj. | 48 subj. | 48 subj. | 48 subj. |
| Fixed Effects | yes | no | yes | no | yes | no |
| Adj. R-squared [0,1] | .431 | | .324 | | .361 | |
| N | 1,200 | 1,200 | 864 | 864 | 768 | 768 |

*Notes:* The regressions show for each treatment how the 'Decisions in Situation A' depend on the 'Efficiency factor in Sit. A' (which is the potential donation amount divided by 100 points) and on the cheapness of Sit. A (captured in the dummy 'Donation in Sit. A unlikely' which is 1 if donation in Sit. A was impossible with $p = 0.9$ and 0 if it was impossible with $p = 0.1$). The 'Decision in Situation A' is a binary variable which has value 1 if the decision-maker decided for donation and 0 if he or she decided to take the 100 points.
***(**/*) depict significance at the 1 (5/10) percent level.

comes weaker - and not stronger as hypothesized in hypothesis **2** - when moving from $T1$ to $T2$ to $T3$. To validate this conjecture, we performed an OLS and logit regression (Table 3 and 4). In Table 3 the picture-treatment ($T2$) serves as a baseline. The impact of cheapened signaling costs in $T2$ is captured in the variable 'Donation in Sit. A unlikely'. The first interaction term ('Donation in Sit. A unlikely × T1') tells us that this effect is slightly lower than in $T1$, yet not significantly so. The second interaction term ('Donation in Sit. A unlikely × T3') indicates that the effect is higher than in $T3$ - yet, again not significantly so. Thus, the effect of $p$ in $T2$ is not significantly different from the effect in $T1$ or $T3$. However, Table 4 shows that increasing the probability that donation is impossible had a significantly different effect for $T1$ than for $T3$. When signaling became cheaper subjects increased their donation decisions more in $T1$ than in $T3$ (OLS: $p < 0.1$, logit: $p < 0.05$).

**Result 2:** *The effect of cheapened signaling costs is not significantly stronger for treatments with higher reputational incentives. On the contrary, the effect is significantly weaker in the picture+bonus-treatment than in the anonymous-treatment.*

By now we found that cheapening signaling costs can in some settings affect how individuals decide. However, it should also influence how others evaluate these decisions. After all, donation decisions that are implemented almost never should raise doubts about whether high altruism (or the cheapness of the situation) was the dominant motive behind

Table 3: Decisions in Situation A (T2 as baseline)

| | OLS | Logit |
|---|---|---|
| Efficiency factor in Sit.A | 0.20*** | 0.93*** |
| | (0.02) | (0.09) |
| Donation in Sit.A unlikely (p=0.9) | 0.06* | 0.30* |
| | (0.04) | (0.18) |
| T1: anonymous | -0.10* | -0.47* |
| | (0.05) | (0.27) |
| T3: with picture+bonus | 0.09 | 0.45 |
| | (0.06) | (0.27) |
| Donation in Sit.A unlikely × T1 | 0.02 | 0.12 |
| | (0.05) | (0.24) |
| Donation in Sit.A unlikely × T3 | -0.08 | -0.36 |
| | (0.06) | (0.26) |
| Constant | 0.06 | -2.00*** |
| | (0.04) | (0.23) |
| Adj. R-squared [0,1] | .148 | |
| Clusters | 165 subj. | 165 subj. |
| N | 2,832 | 2,832 |

*Notes:* The regressions show how 'Decisions in Situation A' are affected by the 'Efficiency factor in Sit. A', the cheapness of Sit. A (captured in 'Donation in Sit.A unlikely'), the reputational incentive in Sit.A (captured in the two treatment dummies 'T1: anonymous' and 'T3: with picture + bonus') and interactions between the reputational incentive and cheapness in Sit.A ('Donation in Sit. A unlikley × T1' and 'Donation in Sit. A unlikely × T2'). The 'Decision in Situation A', 'Efficiency factor in Sit. A' and 'Donation in Sit.A unlikely' are defined as in Table 2.
***(**/*) depict significance at the 1 (5/10) percent level. Standard errors are clustered on subject-level and shown in parentheses.

the decision. Observers should understand that one does not have to be extraordinarily altruistic to commit to a donation which most likely cannot be done anyway. Consequently, the expected altruism and the estimates should be lower when donations are likely impossible (see Hypothesis **3**). We will now test whether observers' estimates actually follow this principle.

Recall that observers' estimates are the incentivized belief that a person donated in *Situation B* given his or her decision in *Situation A*. Observers were first informed about (i) how likely donation was impossible for the observed person in *Situation A* and (ii) how much the observed person could donate in *Situation A and B* if donation was possible. Then, they had to make two estimates: one assuming that the observed person had de-

Table 4: Decisions in Situation A (T1 as baseline)

| | OLS | Logit |
|---|---|---|
| Efficiency factor in Sit.A | 0.20*** | 0.93*** |
| | (0.02) | (0.09) |
| Donation in Sit.A unlikely (p=0.9) | 0.08*** | 0.42*** |
| | (0.03) | (0.15) |
| T2: with picture | 0.10* | 0.47* |
| | (0.05) | (0.27) |
| T3: with picture+bonus | 0.19*** | 0.91*** |
| | (0.05) | (0.26) |
| Donation unlikely × T2 | -0.02 | -0.12 |
| | (0.05) | (0.24) |
| Donation unlikely × T3 | -0.10* | -0.48** |
| | (0.05) | (0.24) |
| Constant | -0.03 | -2.47*** |
| | (0.04) | (0.24) |
| Adj. R-squared [0,1] | .148 | |
| Clusters | 165 subj. | 165 subj. |
| N | 2,832 | 2,832 |

*Notes:* These are essentially the same regressions as in Table 3 except of that we used the picture-treatment, T2, (and not the anonymous-treatment, T1) as baseline. ***(**/*) depict significance at the 1 (5/10) percent level. Standard errors are clustered on subject-level and shown in parentheses.

cided for donation in *Situation A* and one assuming that the observed person had decided to keep the 100 points in *Situation A*.

Table 5 and 6 show how the efficiency factor in *Situation A*, the efficiency factor in *Situation B* and the probability that donation is impossible in *Situation A* influence observers' estimates for a donation or keep-decision in *Situation A*, respectively. To unravel potential differences across treatments, we again treat each reputational treatment as separate unit. This also allows us to include subject fixed effects into the regressions because except of the reputational incentive, which is varied between treatments, all other determinants are varied within subjects. This regression design rules out all differences in estimation behavior between subjects and casually identifies the effects of the manipulation of $p$ and the efficiency factor $X$ in *Situation A and B*.

For the regressions in Table 5 and 6 we excluded observations in which observers estimated about 50 for both *Situation A*-decisions because those estimates most likely reflect the desire to get a certain payoff and not the observer's true belief.[13] Dropping these

[13]Due to the incentive structure of the quadratic scoring rule subjects could secure themselves a certain

observations does not fundamentally change the results (see Table A3 and A4 in the Appendix for the complete dataset).

Table 5: Estimates given a donation decision in Situation A

| | anonymous | | picture | | picture+bonus | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Efficiency factor | -1.34 | -2.41 | -1.50 | -0.53 | -6.04** | -5.82** |
| in Sit.A | (1.14) | (1.31) | (1.76) | (1.44) | (1.86) | (1.72) |
| Efficiency factor | 6.37*** | 6.83*** | 7.97** | 7.74** | 6.65** | 5.90** |
| in Sit.B | (1.63) | (1.91) | (2.92) | (2.99) | (1.86) | (1.84) |
| Donation in Sit.A | -0.47 | -0.44 | -4.49* | -4.18** | 2.26 | 2.96 |
| unlikely (p=0.9) | (2.03) | (2.32) | (1.87) | (1.43) | (2.08) | (2.18) |
| Model | OLS | OLS | OLS | OLS | OLS | OLS |
| Clusters | 9 MG | 9 MG | 6 MG | 6 MG | 6 MG | 6 MG |
| Fixed effects | yes | no | yes | no | yes | no |
| Adj. R-squared [0,1] | .456 | .0424 | .337 | .0611 | .399 | .0561 |
| N | 906 | 906 | 619 | 619 | 540 | 540 |

*Notes:* The dependent variable is the 'Estimate given a donation decision in Situation A', i.e. the observer's belief that the observed person donated in Situation B assuming that this person decided for donation in Situation A. The independent variables are the observed specifics of Sit.A and B: the 'Efficiency factor in Sit. A', the 'Efficiency factor in Sit.B' and the cheapness of Situation A ('Donation in Sit.A unlikely (p=0.9)').
***(**/*) depict significance at the 1 (5/10) percent level. Standard errors are clustered at the matching group (MG) level. Observations for which both estimates of a round are about 50 are excluded; for whole sample see Table A3 in Appendix.

To evaluate the empirical validity of hypothesis **3** we start with estimates made for the scenario in which the observed person took the 100 points in *Situation A* (Table 6). The regressions show a mixed picture. For the anonymous-treatment ($T1$) and the picture-treatment ($T2$) the coefficient of the dummy which indicates that the observed donation decisions were unlikely realized is directionally as expected, i.e. negative. However, in $T1$ it is insignificant and in $T2$ only weakly significant. For the picture+bonus-treatment we observe a positive, albeit insignificant, coefficient. Overall, this offers little support for the hypothesis that observers discount the diagnostic value of donation decisions if deciding for donation is cheap.

payoff of 75 points by setting both estimates of a round, i.e. the one given a donation and the one given a keep-decision, on 50. A sufficiently risk-averse subject will thus center his estimates around 50 independent of their true beliefs/expectations. Since estimates had to be entered via sliders and it was very tedious to hit exactly 50, we excluded all estimates that fell into the interval $[45, 55]$. These are about 27% of the estimates.

Table 6: Estimates given a keep decision in *Situation A*

| | anonymous | | picture | | picture+bonus | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Efficiency factor | -2.14 | -1.60 | -2.07 | -1.18 | -5.75** | -5.48** |
| in Sit.A | (1.44) | (1.74) | (1.78) | (1.97) | (1.49) | (1.37) |
| Efficiency factor | 4.15** | 3.31** | 4.89 | 2.46 | 7.10** | 6.30** |
| in Sit.B | (1.64) | (1.38) | (3.05) | (2.54) | (2.09) | (1.96) |
| Donation in Sit.A | 0.81 | 0.29 | 3.24* | 3.54** | -2.75 | -0.94 |
| unlikely (p=0.9) | (1.67) | (1.93) | (1.34) | (1.37) | (2.72) | (2.87) |
| Model | OLS | OLS | OLS | OLS | OLS | OLS |
| Clusters | 9 MG | 9 MG | 6 MG | 6 MG | 6 MG | 6 MG |
| Fixed effects | yes | no | yes | no | yes | no |
| Adj. R-squared [0,1] | .548 | .00952 | .473 | .00454 | .47 | .0556 |
| N | 906 | 906 | 619 | 619 | 540 | 540 |

*Notes:* The dependent variable is the 'Estimate given a keep decision in Situation A', i.e. the observer's belief that the observed person donated in Situation B assuming that this person decided to keep the 100 points in Situation A. The independent variables are the observed specifics of Sit.A and B: the 'Efficiency factor in Sit. A', the 'Efficiency factor in Sit.B' and the cheapness of Situation A ('Donation in Sit.A unlikely (p=0.9)').
***(**/*) depict significance at the 1 (5/10) percent level. Standard errors are clustered at the matching group (MG) level. Observations for which both estimates of a round are about 50 are excluded; for whole sample see Table A3 in Appendix.

This conclusion is strengthened by the estimates which observers made for the scenario in which the observed person took the 100 points in *Situation A* (Table 6). The coefficients of the anonymous- and picture+bonus-treatment ($T1$ and $T3$) are again insignificant. Only in the picture-treatment ($T2$) the information that donation decisions in *Situation A* were only very unlikely implemented, significantly shifted estimates. However, subjects entered higher (and not lower) estimates if a person did not donate in *Situation A* and donation was very likely impossible. This is surprising because our model suggests that keeping the money becomes a stronger signal for low altruism when deciding for donation becomes cheaper. When signaling altruism gets cheaper, the more altruistic individuals among the individuals who keep the money, switch to a donation decision. Therefore, the group of individuals who still keep the money is on average less altruistic.

**Result 3:** *Overall, observers' estimates seem not to be significantly influenced by the cheapness of signaling. The probability that donation was impossible has only in the* *picture-treatment a weakly significant effect - yet, the effect is opposing to the prediction if observers evaluated keep-decisions.*

## 5.2 The effect of other factors

We complete the analysis of perceptions by discussing how the other situational factors determined observers' estimates. Clearly, the factor that drove estimations most strongly is the efficiency factor in *Situation B.* In all treatments observers assigned a significantly higher probability to the event that the observed person donated in *Situation B* if waiving the 100 points resulted in a larger donation amount. On contrary to this, the efficiency factor in *Situation A* affected estimates differently across treatments. Intuitively, donation decisions in *Situation A* should be less predictive for high altruism if the donation amount in *Situation A* is high because then also less altruistic persons find it worthwhile to decide for donation. Following this reasoning a high 'efficiency factor in Sit.A' should decrease estimates (hypothesis **4.1a**). This holds, however, only in the treatment with the picture and bonus, $T3$. While also in the other treatments the coefficients go into the expected direction, they do not reach significance (Table 5 and 6).

**Result 4.1a:** *Observers' estimates (i.e. the incentivized beliefs that the observed person donated in Situation B) only significantly decreased in the efficiency factor of Situation A if the observed person received the estimate as monetary bonus.*

Seemingly, observers only paid significant attention to the efficiency factor in *Situation A* if they knew that the person, who they evaluated, will be paid their estimate. A possible explanation for this could be that only in $T3$ the incentive to guess a person's (true) altruism was strong enough. In $T3$ people could use high estimates to reward others. If observers wanted to reward altruism (but not selfishness), they would have an additional incentive to evaluate the altruism of others correctly. Consequently, they might want to pay more attention to the efficiency factor of *Situation A* because this helps to assess how diagnostic the observed decisions are. While theoretically also the probability that donation is impossible $p$ should matter, empirically it does not (see result 1). The insignificance of the coefficient 'Donation in Sit.A unlikely (p=0.9)' is therefore not at odds with this possible explanation.

Figure 3 displays the cumulative distribution functions of estimates for a given efficiency factor in *Situation B.* An inspection of the two graphs shows that for all efficiency factors there is a higher concentration on low estimates if the observed person kept the 100 points in *Situation A* than if he or she decided for donation in *Situation A*. This observation is also supported by the estimates that participants made on average. If a person decided for donation in *Situation A* the average estimate was 46.15%, while it was with 32.58% significantly lower if a person decided against donation (paired t-test, $p < 0.0001$).

This is in line with hypothesis **4.1b**.

**Result 4.1b:** *Observers' estimates (i.e. the incentivized belief that the observed person donated in Situation B) was significantly higher when the observed person donated in Situation A.*

We now return to donation decisions evaluate hypothesis **4.2a** and **4.2b**. Figure 4 graphically illustrates how strongly decisions were influenced by the efficiency factor in the respective situation. Donation decisions more than triple when moving from the lowest donation amount to the highest: If waiving 100 points resulted in a (potential) donation of 300 points, we observed in about half of the decision-problems donation decisions (66% in *Situation A* and 48% in *Situation B*), while subjects decided in less than a fifth of the decision-problems for donation if waiving 100 points resulted only in a potential donation of 50 points (19% donation decisions in *Situation A* and 11% in *Situation B*). The regressions in Table 2, 3 and A1, A2 (Appendix) confirm that the efficiency factor significantly influenced decisions.

**Result 4.2a:** *In Situation A and B more individuals decided for donation if donation was more efficient in the respective situation.*

Moreover, figure 5 shows that subjects, who took part in treatments with a stronger reputational incentive, decided more often for donation in *Situation A* (as hypothesized in **4.2a**). Presenting *Situation A*-decisions together with a picture of the decision-maker increased donation decisions by 9 percentage points (comparison $T1$ and $T2$), whereas adding a bonus for pro-social appearance lead to a further increase of 4 percentage points (comparison $T2$ and $T3$). However, only the first increase is significant ($p < 0.1$, see Table 3).

**Result 4.2b:** *Strengthening the reputational incentives by adding a picture lead to more donation decisions in Situation A.*

While it is very intuitive that the treatment affected decisions in *Situation A*, figure 5 suggests that it also affected decisions in *Situation B*. This is less intuitive because $T1$, $T2$ and $T3$ varied only the reputational incentive for *Situation A* but not for *Situation B*. However, a regression analysis shows that, after controlling for *Situation A*-decisions, the treatment had no significant effect on *Situation B*-decisions (Table A1 and A2 in the Appendix). Since the coefficient of 'Donation Decision in Sit.A' is highly significant, we suspect that (some) subjects had a desire to be consistent between *Situation A and B*. Such a desire would make sense if they wanted to prove themselves that the changed reputational incentive does not change their decisions much because their decisions are mainly driven by altruism and not by image-concerns or the bonus. That is, if subjects

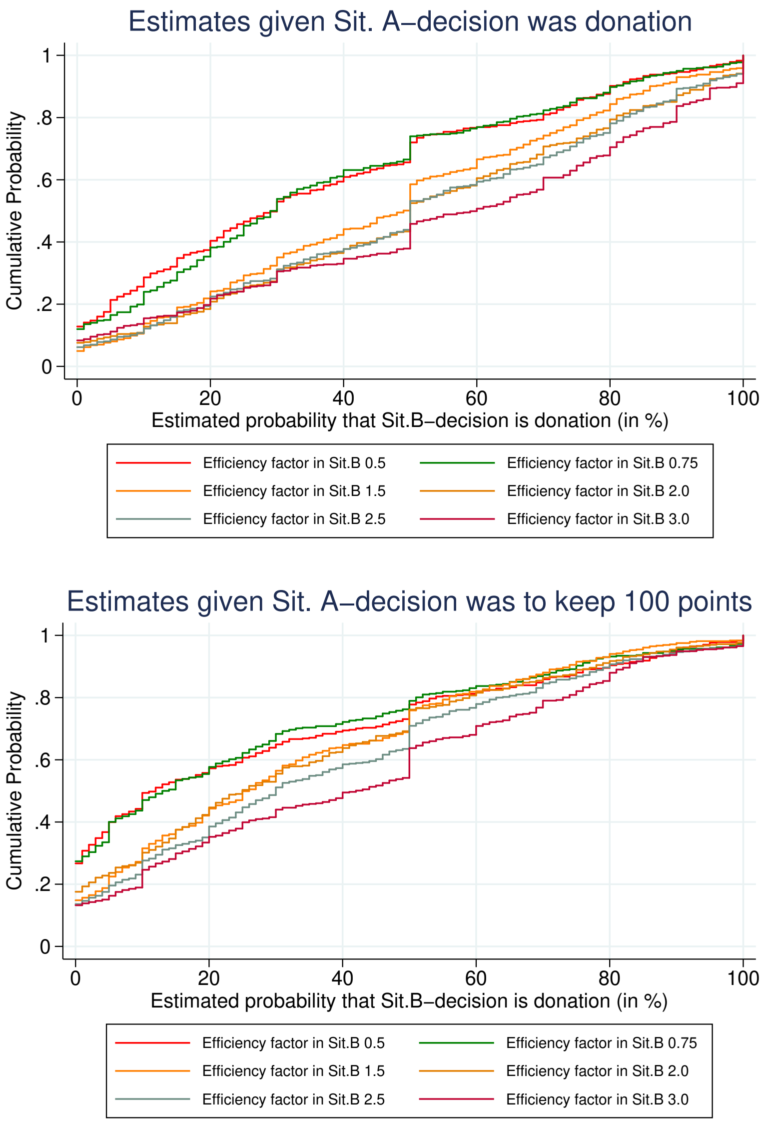



Figure 3: This figure shows how likely observers believed that a decision-maker donated in Sit.B given that this decision-maker decided for donation in Sit.A (upper graph) or given that this person decided to keep the 100 points in Sit.A (lower graph). The beliefs are reflected in the probability estimate on the x-axis (in %). The y-axis shows the cumulative frequency of each probability estimate (separately for each efficiency factor in Sit.B).

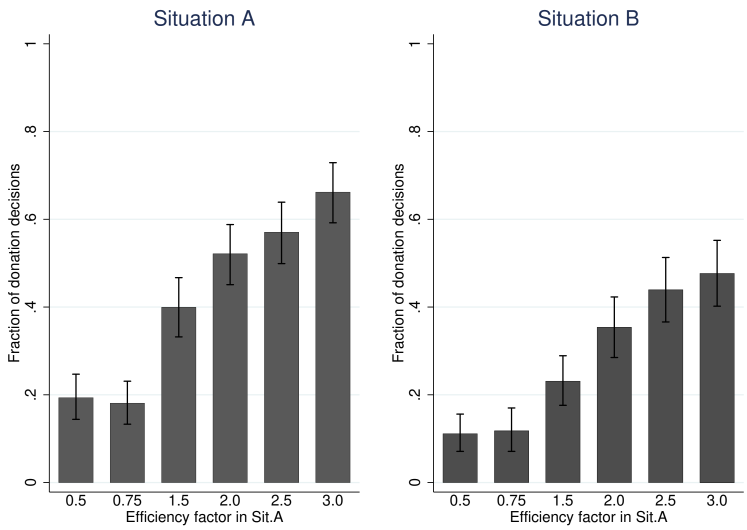



Figure 4: The figure shows the share of donation decisions in *Situation A* (left) and *Situation B* (right) across donation amounts and their 95% confidence intervals (based on a bootstrap with 10,000 repetitions and clustering at subject level). The efficiency factor multiplied by 100 gives the donation amount. The donation amount specifies how many points the decision-maker could (potentially) donate if he decided for donation instead of taking 100 points.

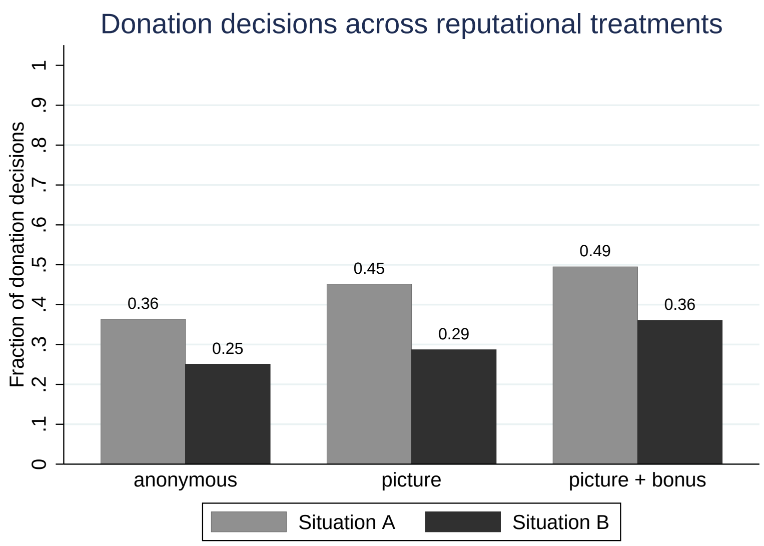



Figure 5: Decisions in *Situation A* and *Situation B* depending on the reputational treatment. The x-axis depicts the frequency of donation decisions; the y-axis shows whether *Situation A*-decisions were revealed anonymously (T1), with picture (T2) or with picture and there was additionally a monetary bonus for a pro-social appearance (T3).

not only had social-image concerns but also self-image concerns.

## 6 Conclusion

This paper studies the tendency to decide pro-socially and the evaluations of these decisions in the presence of cheapened signaling costs. We find that the effect of cheapened signaling costs crucially depends on the reputational incentive: when costly donation decisions were more likely infeasible (i.e. signaling cheaper), more decision-makers decided for donation - but only if the reputational incentive was not too strong, suggesting a crowding-out effect. Transferring this result to situations outside of the lab, it might help to promote pro-social decisions. In many cheap signaling scenarios people only have vague ideas about how likely they have to fulfill their costly promise. If there is a reason to believe that this likelihood is biased upwards and reputational incentives are low, clarifying how unlikely the costs are realized could foster pro-social decisions.

We also find that expected signaling costs affected the evaluations of observers about the true level of altruism in a biased way. In situations with strong non-monetary reputation, observing a donation is less telling about a high level of altruism if the expected signaling costs become cheaper. After deciding against donation in case the expected signaling costs are lower should in theory also decreases the perceptions. In the data, however, we find an increase in the perceptions. For future research it might be interesting to test the robustness of this unintuitive deviation and, in case it holds in different settings, where it sources from.

## 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.

Bénabou, R., A. Falk, and J. Tirole (2018a). Eliciting Moral Preferences.

Bénabou, R., A. Falk, and J. Tirole (2018b). Narratives, Imperatives, and Moral Reasoning. *NBER Working Paper Series* (249429), 52.

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

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.

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

Falk, A. and N. Szech (2017). Diffusion of Being Pivotal and Immoral Outcomes. *Working Paper*, 1–28.

Falk, A. and J. Tirole (2016). Narratives , Imperatives and Moral Reasoning. *Working Paper* (249429), 1–50.

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.

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.

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., A. K. Wagner, B. Grofman, and S. Voigt (2016). Experimental Evidence on Expressive Voting Experimental Evidence on Expressive Voting *. pp. 1–22.

# A Appendix

## A1. Proofs

**Proof for Proposition 4.**

**4.1a**

An increase in $X$ decreases $\bar{\mu}$ (see 4.2b) and thereby also $\frac{\bar{\mu}+1}{2}$ and $\frac{0+\bar{\mu}}{2}$.

**4.1b**

The perceived altruism is $E[\mu_i|\mu_i \geq \bar{\mu}] = \frac{\bar{\mu}+1}{2}$ if the person decided for donation and $E[\mu_i|\mu_i < \bar{\mu}] = \frac{0+\bar{\mu}}{2}$ otherwise. Since $\bar{\mu} \geq 0 : \frac{\bar{\mu}+1}{2} > \frac{0+\bar{\mu}}{2}$.

**4.2a**

Note that $\mu_i \sim U[0,1]$ implies $\bar{\mu} \geq 0$ and as we focus on separating EQ, in which at least some types do not donate, $\bar{\mu} > 0$ must hold. From $\bar{\mu} > 0 \Leftrightarrow \frac{1}{X} \cdot \left(1 - \frac{1}{200} \cdot \frac{\lambda}{(1-p)}\right) > 0$ it follows that $\lambda < 200(1-p)$.

Since $\lambda < 200(1-p)$ holds, it holds that

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

**4.2b**

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

## A2. Further Results

Table A1: Decisions in Situation B (T1 as baseline)

| | OLS | Logit |
|---|---|---|
| Efficiency factor in Sit.B | 0.16*** | 0.93*** |
| | (0.02) | (0.10) |
| Donation Decision in Sit.A | 0.22*** | 1.26*** |
| | (0.03) | (0.19) |
| Efficiency factor in Sit.A | -0.07*** | -0.41*** |
| | (0.01) | (0.08) |
| Donation in Sit.A unlikely (p=0.9) | 0.01 | 0.08 |
| | (0.01) | (0.08) |
| T2: with picture | 0.02 | 0.11 |
| | (0.04) | (0.25) |
| T3: with picture+bonus | 0.08 | 0.43 |
| | (0.05) | (0.28) |
| Constant | 0.00 | -2.72*** |
| | (0.03) | (0.25) |
| Adj. R-squared [0,1] | .158 | |
| Clusters | 165 subj. | 165 subj. |
| N | 2,832 | 2,832 |

*Notes:* These regressions show how the 'Decision in Situation B' (which is 1 if the decision in Sit.B was donation and 0 otherwise) depends on the 'Efficiency factor in Sit.B' (i.e. the donation amount in Sit.B divided by 100 points), the subject's decision in Situation A of the same round ('Donation Decision in Sit.A' which is 1 if the subject decided for donation in Sit.A and 0 otherwise), the 'Efficiency factor in Sit.A'(i.e. the donation amount in Sit.A divided by 100 points), the probability that donation in Sit.A was impossible ('Donation in Sit.A unlikely (0.9)') and the treatment (the dummy 'T2: with picture' is 1 for $T2$ and otherwise 0; the dummy 'T3: with picture + bonus' is 1 for $T3$ and 0 otherwise; $T1$ serves as baseline).
***(**/*) depict significance at the 1 (5/10) percent level. Standard errors are clustered on subject-level and shown in parentheses.

Table A2: Decisions in Situation B (T2 as baseline)

| | OLS | Logit |
|---|---|---|
| Efficiency factor in Sit.B | 0.16*** | 0.93*** |
| | (0.02) | (0.10) |
| Donation Decision in Sit.A | 0.22*** | 1.26*** |
| | (0.03) | (0.19) |
| Efficiency factor in Sit.A | -0.07*** | -0.41*** |
| | (0.01) | (0.08) |
| Donation in Sit.A unlikely (p=0.1) | 0.01 | 0.08 |
| | (0.01) | (0.08) |
| T1: anonymous | -0.02 | -0.11 |
| | (0.04) | (0.25) |
| T3:with picture+bonus | 0.06 | 0.33 |
| | (0.05) | (0.28) |
| Constant | 0.02 | -2.61*** |
| | (0.03) | (0.24) |
| Adj. R-squared [0,1] | .158 | |
| Clusters | 165 subj. | 165 subj. |
| N | 2,832 | 2,832 |

*Notes:* Same regressions as in Table A1 except of using $T2$ as baseline.
***(**/*) depict significance at the 1 (5/10) percent level. Standard errors are clustered on subject-level and shown in parentheses.

Table A3: Estimates given a donation decision in Sit.A (all observations)

| | anonymous | | picture | | picture+bonus | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Efficiency factor | -1.69 | -2.27 | -1.18 | -0.01 | -4.36** | -3.97 |
| in Sit.A | (1.03) | (1.25) | (2.17) | (1.95) | (1.65) | (2.17) |
| Efficiency factor | 8.42*** | 8.34*** | 8.70** | 8.69** | 6.51** | 6.88** |
| in Sit.B | (1.86) | (2.10) | (2.73) | (2.53) | (1.98) | (1.77) |
| Donation in Sit.A | -0.90 | -0.94 | -1.86 | -1.84 | 0.49 | 0.51 |
| unlikely (p=0.9) | (1.31) | (1.30) | (1.26) | (1.18) | (1.20) | (1.16) |
| Model | OLS | OLS | OLS | OLS | OLS | OLS |
| Clusters | 9 MG | 9 MG | 6 MG | 6 MG | 6 MG | 6 MG |
| Fixed effects | yes | no | yes | no | yes | no |
| Adj. R-squared [0,1] | .506 | .058 | .33 | .0697 | .501 | .0435 |
| N | 1,200 | 1,200 | 864 | 864 | 768 | 768 |

*Notes:* For this Table we performed the same regressions as in Table 5, yet using all observations (i.e. also those for which both estimates in a round were centered around 50). The dependent variable is the 'Estimate given a donation decision in Situation A', i.e. the observer's belief that the observed person donated in Situation B assuming that this person decided for donation in Situation A. The independent variables are the observed specifics of Sit.A and B: the 'Efficiency factor in Sit. A', the 'Efficiency factor in Sit.B' and the cheapness of Situation A ('Donation in Sit.A unlikely (p=0.9)').
***(**/*) depict significance at the 1 (5/10) percent level. MG denotes matching groups.

Table A4: Estimates given a keep decision in Sit.A (all observations)

| | anonymous | | picture | | picture+bonus | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Efficiency factor in Sit. A | -1.54 | -1.62 | -1.96 | -0.60 | -4.76*** | -4.12* |
| | (1.02) | (1.33) | (1.90) | (2.19) | (0.92) | (1.78) |
| Efficiency factor in Sit.B | 5.37*** | 4.79*** | 5.54 | 5.02 | 7.45** | 7.11** |
| | (1.33) | (1.29) | (3.01) | (2.81) | (1.93) | (2.01) |
| Donation in Sit.A unlikely (p=0.9) | 0.11 | 0.09 | 3.51* | 3.52* | -2.49 | -2.49 |
| | (1.74) | (1.71) | (1.50) | (1.44) | (2.43) | (2.32) |
| Model | OLS | OLS | OLS | OLS | OLS | OLS |
| Clusters | 9 MG | 9 MG | 6 MG | 6 MG | 6 MG | 6 MG |
| Fixed effects | yes | no | yes | no | yes | no |
| Adj. R-squared [0,1] | .546 | .0236 | .435 | .0224 | .494 | .0531 |
| N | 1,200 | 1,200 | 864 | 864 | 768 | 768 |

*Notes:* For this Table we performed the same regressions as in Table 6, yet using all observations (i.e. also those for which both estimates in a round were centered around 50). The dependent variable is the 'Estimate given a keep decision in Situation A', i.e. the observer's belief that the observed person donated in Situation B assuming that this person decided to keep the 100 points in Situation A. The independent variables are the observed specifics of Sit.A and B: the 'Efficiency factor in Sit. A', the 'Efficiency factor in Sit.B' and the cheapness of Situation A ('Donation in Sit.A unlikely (p=0.9)').
***(**/*) depict significance at the 1 (5/10) percent level. MG denotes matching groups.

## A3. Decision Screens

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

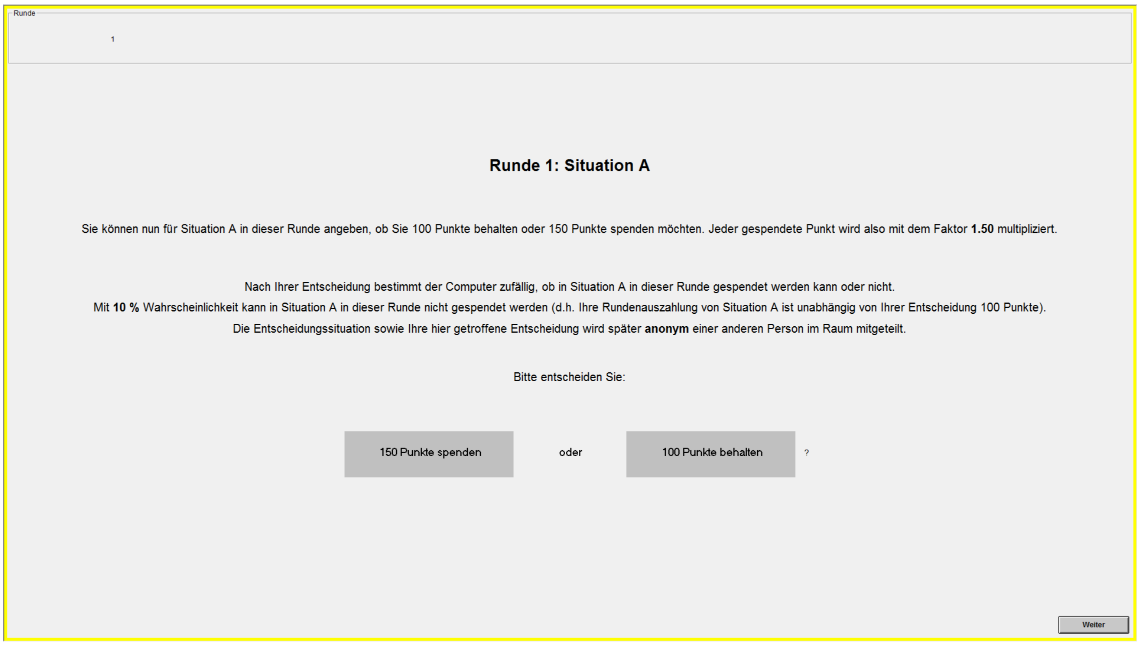



Figure A1: Treatment T1, Situation A.

*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."

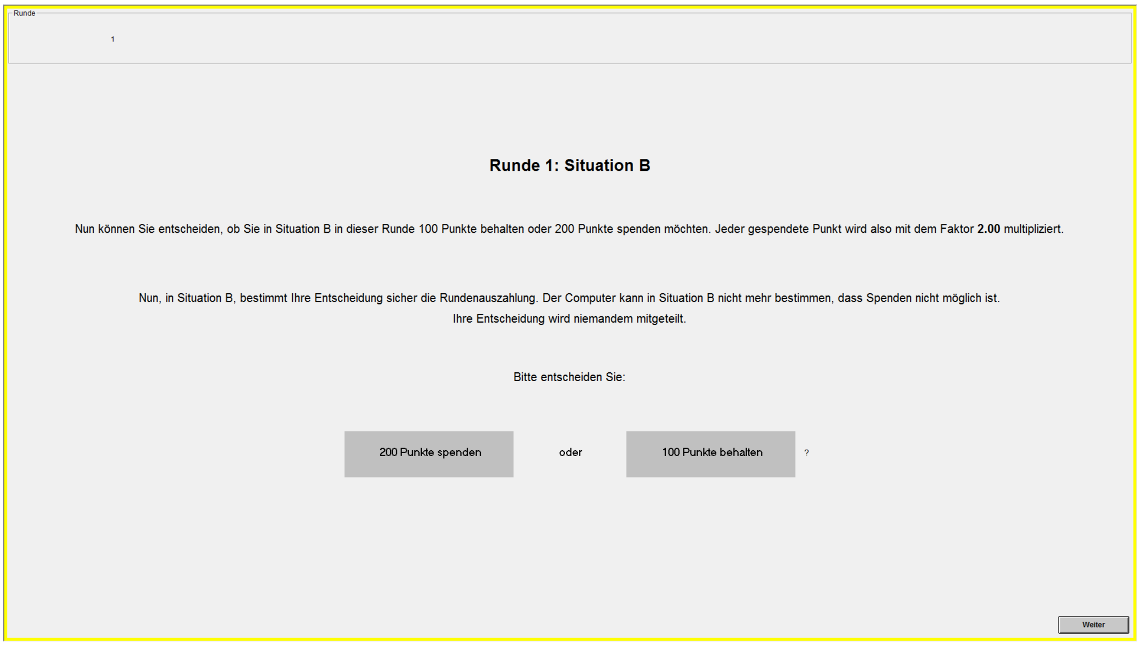



Figure A2: All Treatments, Situation B.

*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 roundpayoff. The computer cannot decide donation to be impossible in situation B. Your decision is not communicated to anyone. Please decide: Donating 200 points or keeping 100 points?"

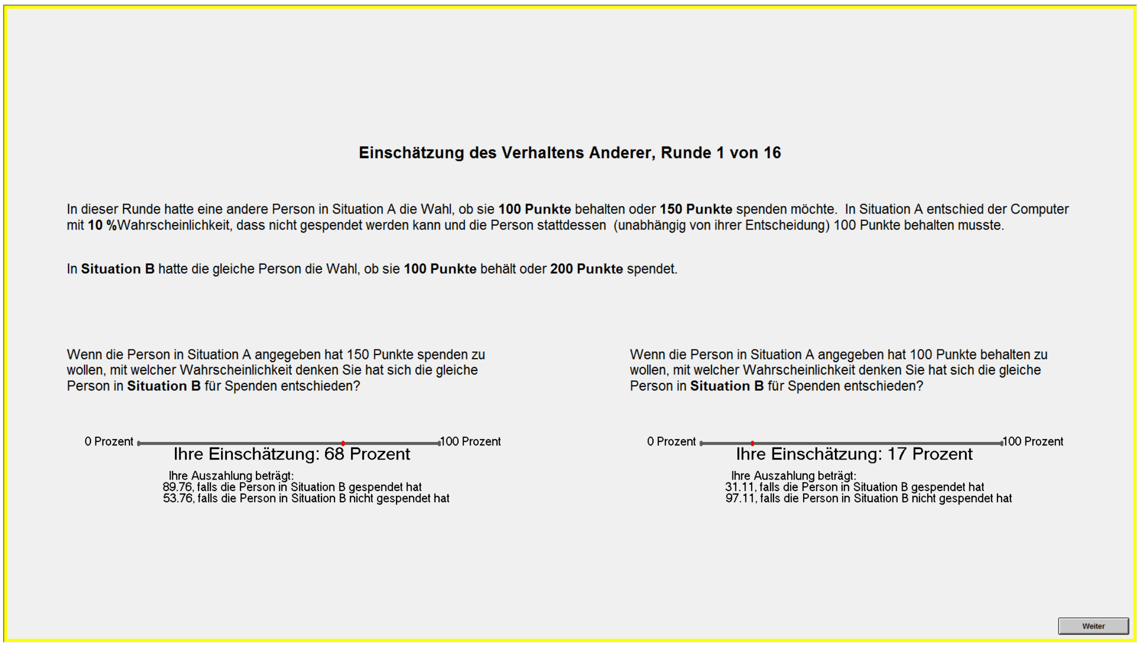



Figure A3: All treatments, Observer Screen.

*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 the person (independent on the decision) had to keep 100 points. In situation B, the same person had to choice between keeping 100 points or donating 200 points. If the person indicated to donate 150 points, 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, 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 A4: 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 in this case was 10%."

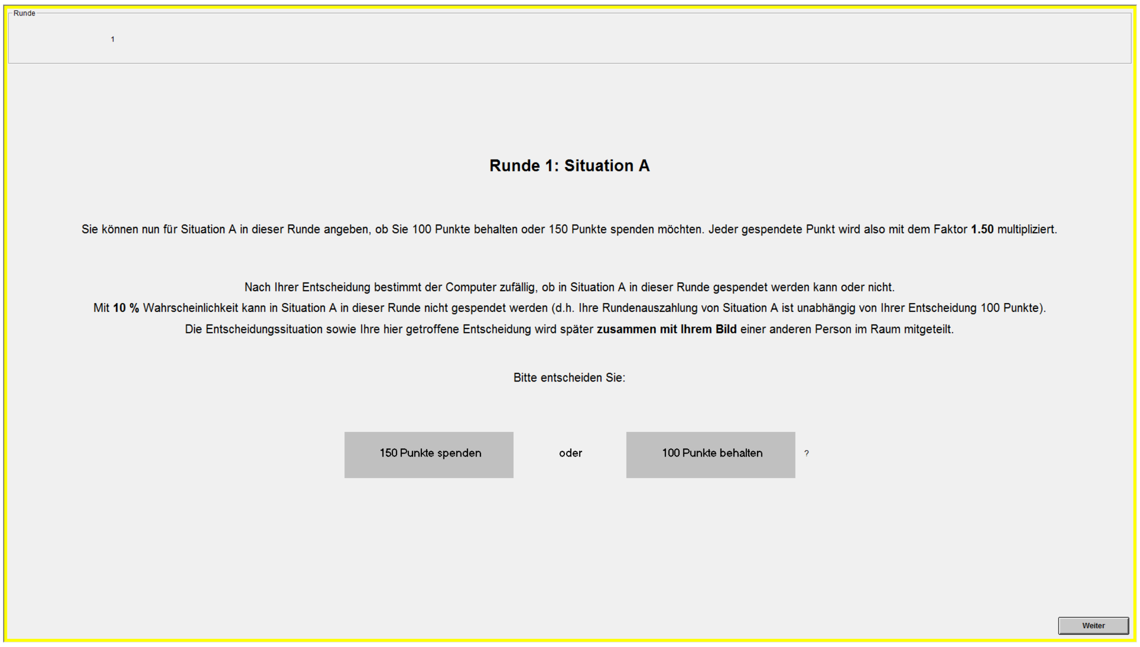



Figure A5: Treatment T2, Situation A.

*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 together with your picture to another person in the room. Please decide: donate 150 points or keep 100 points."

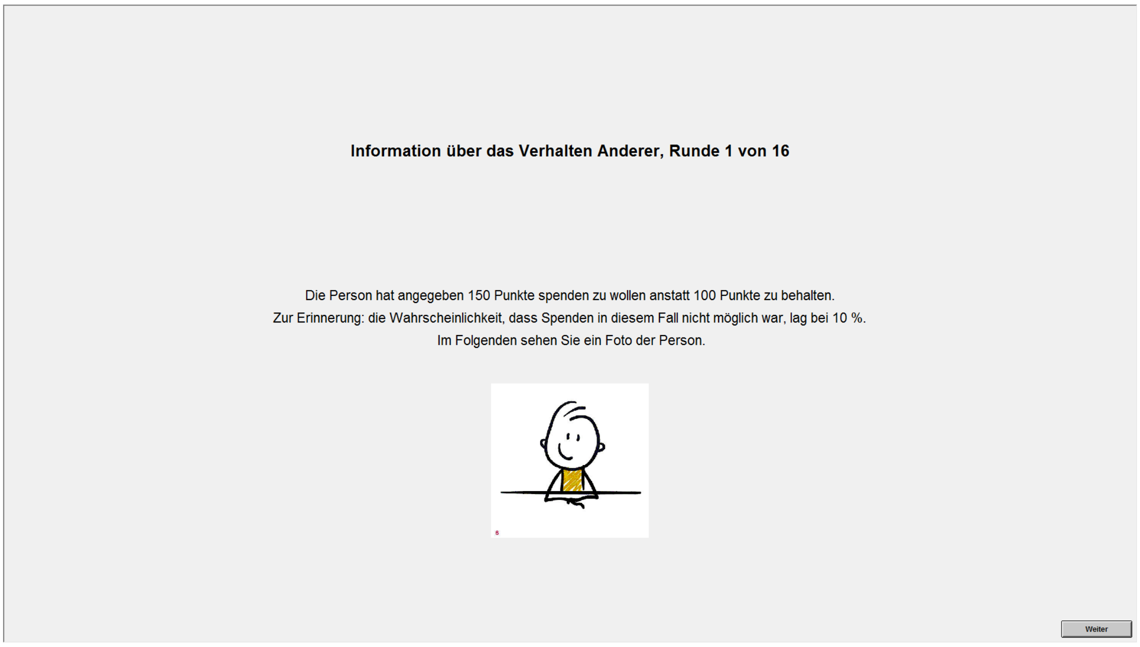



Figure A6: Treatment T2 and T3, Feedback of Observer.

*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 in this case was 10 %. In the following, you can see a picture of this person."

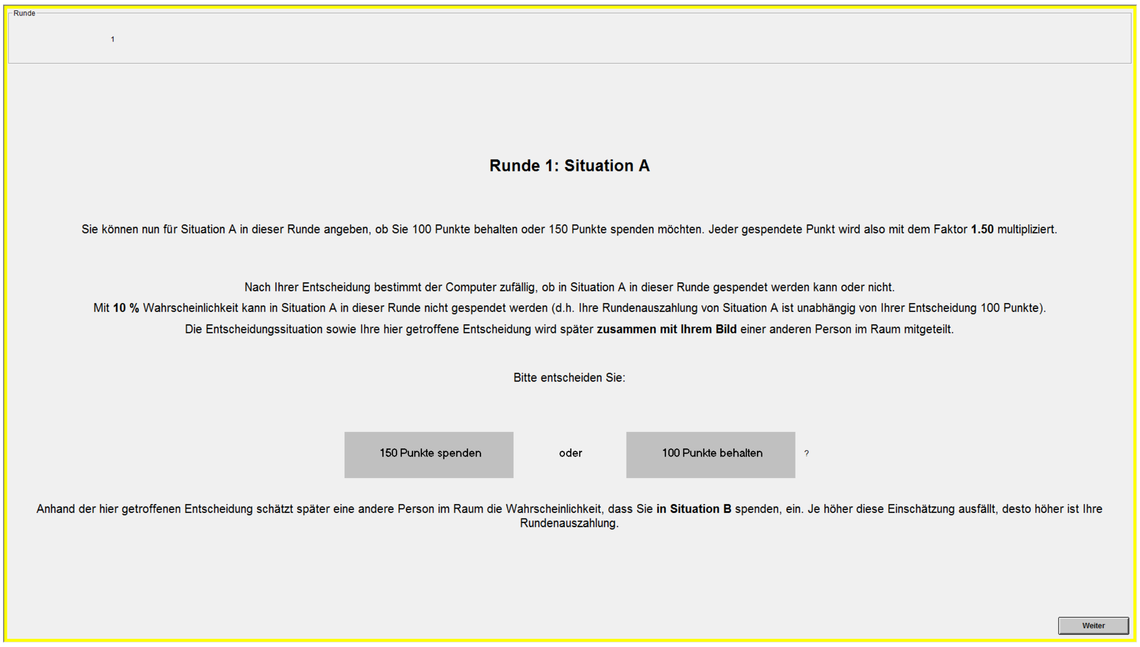



Figure A7: Treatment T3, Situation A.

*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 1.5. 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 together with your picture to another person in the room. Please decide: donate 150 points or keep 100 points. According to your decision here, another person in the room estimates the probability that you decide in Situation B. The higher the estimate, the higher your round-payoff."