I Care What You Think: Social Image Concerns and the Strategic Revelation of Past Pro-Social Behavior

Ferdinand A. von Siemens

CESifo

I Care What You Think:
Social Image Concerns and the
Strategic Revelation of Past Pro-Social Behavior

Abstract

This article studies whether people want to control which information on their own past pro-social behavior is revealed to other people. Participants in an experiment are assigned a color which depends on their own past pro-sociality. They can then spend money to increase or decrease the probability with which their color is revealed to another participant. The data show that participants are more likely to reveal colors that have a more favorable informational content. This pattern is not found in a control treatment in which colors are randomly assigned and thus have no informational content. Regression analysis confirms these findings, also when controlling for the initial pro-social decision. These results complement the existing empirical evidence, and suggests that people strategically manipulate the pro-social impression they make on other people, even though a favorable reputation has no immediate material benefits.

JEL-Codes: C900, D010, D800.

Keywords: social signaling, trust, altruism.

Ferdinand A. von Siemens
Goethe University Frankfurt
Faculty of Economics and Business Administration
Theodor-W.-Adorno-Platz 4
Germany – 60323 Frankfurt
vonsiemens@econ.uni-frankfurt.de

1 Introduction

Field and laboratory experiments suggest that people strategically manipulate and therefore consciously care for their social image, even if a favorable reputation generates no immediate material benefits. Social image concerns indeed seem to influence a wide range of behaviors, like charitable giving, workplace conduct, voting, consumption choices, financial decisions, and investments in education, see Soetevent (2005), Falk & Ichino (2006), Andreoni & Bernheim (2009), Ariely, Bracha & Meier (2009), Mas & Moretti (2009) and further articles cited in the survey by Bursztyn & Jensen (2017). However, almost all existing studies on social image concerns use the same empirical identification strategy: they argue that people care for their social image, because they change their behavior under the scrutiny of a human audience.

Drawing strong conclusion from one dominant experimental paradigm can be dangerous, because findings could be driven by elements of the decision environment that are not related to the considered mechanism. Bursztyn & Jensen (2017) point out that making behavior observable could, for example, interact with social learning or trigger privacy concerns. People might thus change their behavior under the scrutiny of a human audience, although they do not care for their social image. Haley & Fessler (2005), Bateson, Nettle & Roberts (2006), and Burnham & Hare (2007) show that impersonal audiences – the pictures of eyes on a poster or computer screen – promote pro-social behavior. It seems very unlikely that participants want to influence the impression they make on a picture. Subtle facial cues might instead automatically activate an ancient brain system which makes people "feel" watched. People then do care for their social image, but they subconsciously and instinctively respond to false reputation concerns, and they do not consciously and strategically deliberate how their behavior affects their social image. Alternatively, impersonal audiences might influence behavior by increasing self-awareness, see Diener & Wallbom (1976) and Falk (2017). People then adjust their behavior, not because they want to impress others, but because they want to impress themselves.

2

Complementing the pre-dominant experimental paradigm, the present paper thus studies the existence of conscious social image concerns without exogenously varying the observability of behavior. The experiment investigates instead whether people themselves want to vary the observability of their own past behavior – whether they are willing to spend money in order to influence what information on their own past pro-social behavior is revealed to other participants. The experiment therefore tests directly whether people deliberately and consciously control the flow of information and thereby the pro-social impression they make on others.

In the main treatment "Altruism," participants first interact face-to-face with their neighbor. They then make a private donation to a well-known charity. Participants are next assigned a color which depends on their donation as compared to the donations of two randomly determined other participants. They are assigned the color green if they donate more, red if they donate less, and in all other cases they are assigned yellow. Participants finally determine the probability with which their color is revealed to the neighbor with whom they previously interacted. Participants must actively enter a revelation probability, and setting a revelation probability larger or smaller than 50% is costly. They enter their revelation probability privately, and this choice is not directly observed by any other participant. Finally, participants learn the color of their neighbor with the probability determined by the latter.

Colors in Altruism have by design a commonly known informational content: green indicates more pro-social past behavior than yellow, which in turn indicates more pro-social behavior than red. If participants consciously care for their social image, then they should want to strategically manipulate their social image. Sitting isolated in their computer cubicle, they should in cold blood and deliberately set a high revelation probability if assigned green, an intermediate probability if assigned yellow, and a low probability if assigned red. This prediction is very robust, because participants need not speculate on how colors might be interpreted by their neighbors.

Data analysis reveals that participants systematically condition their revelation probabilities on their colors in the main treatment Altruism. Average revelation probabilities are 57% for green, 50% for yellow, and 46% for red. Statistical tests reject that participants do not condition their revelation probabilities on colors. However, this is not enough to conclude that people consciously care for their social image, for two reasons. First, participants might somehow respond to the experimental procedures, and instinctively want to hide colors with common negative connotations. Second, color assignment in Altruism is partly endogenous because it depends on past pro-social behavior. It is possible that due to some unobservable characteristics, participants who behave pro-socially also set a high revelation probability, for example because they want to be transparent. Because they are pro-social, these participants are also assigned more positive colors. Colors and revelation probabilities are then correlated, although participants do not really care for their social image.

The paper pursues two empirical approaches to deal with this potential endogeneity problem. First, behavior in Altruism is compared to behavior in a control treatment "Random". This control treatment Random is identical to Altruism, with the only difference that colors in Random are randomly assigned and therefore reveal no information on past pro-social behavior. If participants adjust their revelation probability because they consciously care for and thus strategically want to manipulate their social image, they should respond to the informational content of the colors. They should thus condition their revelation probabilities on their colors only in Altruism but not in Random. In contrast to the behavior in Altruism, average revelation probabilities in Random are close to 50% for all three colors and statistical tests cannot reject that revelation probabilities do not condition systematically on colors in Random.

Second, regression analysis can try to tackle the potential endogeneity problem by controlling for the pro-social behavior of participants. Controlling for pro-social behavior is possible in the current setup, because assigned colors do not only depend on the donations, but also on the pro-social behavior of the randomly chosen reference groups. Participants with the same donation are consequently sometimes assigned different colors. See also Eil & Rao (2011) for a similar identification approach when looking at the effect of positive and negative feed-back on self-image. The present analysis shows that participants still condition their revelation probabilities on colors in Altruism, both when including the donation decision as linear control variable or when using donation fixed effects.

The present research design is next used to investigate whether social image concerns also affect second-mover behavior in the trust game. Dana, Weber & Kuang (2007) argue that moral wiggle room reduces sharing, because it obfuscates the interpretation of behavior and people care for the impression they make on themselves. Van der Weele, Kulisa, Kosfeld & Friebel (2014) find that moral wiggle room has no effect on return behavior in the trust game, and Fehr & Schneider (2010) show that an impersonal audience does not affect trustworthiness. These studies suggest that self-image or sub-conscious social image concerns have no strong influence on trustworthiness. This raises the question whether conscious social image concerns affect trustworthiness.

To study the importance of conscious social image concerns for trustworthiness, the present paper includes the third treatment "Trust." Participants there are assigned colors conditional on their relative trustworthiness as measured in a trust game. If they care for their social image concerning trustworthiness, they should condition their revelation probabilities on their colors. Average revelation probabilities are 64% for green, 56% for yellow, and 45% for red. Statistical tests reject that people do not condition their revelation probabilities on colors. Regression analysis controlling for initial trustworthiness confirms these results. Participants consciously want to convince others not only of their altruism, but also of their trustworthiness.

The empirical results suggests that people deliberately control and thus consciously care for what others think about them, even in a cold laboratory setup in which a positive social reputation generates no immediate material benefits. The present paper contributes to the existing literature in two ways. First, it complements the existing empirical results on social image concerns, because these are predominantly based on one other experimental paradigm. The results suggest that at least some participants consciously care for the impression they make on strangers, and that they do not only subconsciously respond to false reputation concerns. This insight might be important for developing effective and long-lasting policy interventions, see the discussion by Bursztyn & Jensen (2017).

Second, the present findings also support the foundations of the theoretical literature on social signaling. Austin-Smith & Fryer, Jr. (2005), Bénabou & Tirole (2006) and Ellingsen & Johannesson (2008) argue that social image concerns affect human capital investments and pro-social behavior. The common theoretical interpretation of such signaling models is that

rational players maximize their expected utility while forming correct beliefs about how their behavior is interpreted by their peers. If these models are to explain real world behavior, real people must deliberately consider their strategic options while forming correct beliefs about their peers. The first requirement directly implies that people consciously care for their social image. Moreover, conscious introspection makes it more likely that people form correct beliefs, especially if learning possibilities are limited.[1] The present results shows that at least some people in the experiment consciously contemplate how their behavior affects their social image. This renders the theoretical equilibrium models on social signaling more convincing and more widely applicable.

2   Experimental Design and Procedures

The experiment consists of three parts. The first part of the experiment is designed to increase social proximity between participants so that social image concerns are more likely to affect behavior, see also Dickinson & Villeval (2008) for a similar procedure. Participants are randomly assigned places in the laboratory. They then fill out a questionnaire asking for their favorite color, football club, tree, day, and instrument. The items on the pre-specified lists are chosen so as to have no connection to pro-sociality. Participants cannot earn points by filling out the first questionnaire. They are then given five minutes to discuss the questionnaire – or whatever they want – face-to-face with their neighbor. They sit as pairs in adjacent cubicles with one empty cubicle between each pair. Participants afterwards turn back to their cubicles. From that moment onwards, any further communication between participants is explicitly forbidden. Participants then fill out each other's questionnaire again. They receive 20 points for each answer equal to the initial answer of their neighbor.

In the second part of the experiment, participants have the opportunity to make a pro-social decision. There are three treatments. In the treatments "Altruism" and "Random," participants divide 1000 points in steps of five points between themselves and a well known German children cancer charity (Deutsche Kinderkrebshilfe). The instructions stress that the

---

[1]For example, Mellström & Johannesson (2008) show that monetary incentives reduce the willingness of women to donate blood. Economic theory can explain this behavior as equilibrium outcome of a social signaling game. Women then correctly anticipate that donating blood for money has no positive effect on their social reputation. But where does this beliefs come from? Blood donations are hardly paid in the real world, which is exactly what the theory wants to explain, but which at the same time restricts learning through natural experience or experimentation.

charity is financed exclusively by donations, and they feature a large red slogan emphasizing the importance of donations.

In the treatment "Trust," participants play a trust game in both roles using the strategy method. Both trustor and trustee have an initial endowment of 500 points. The trustor first decides whether to send all or no points to the trustee. In case the trustor decides not to send his points, the game ends, and the trustor and trustee keep their endowments. In case the trustor sends his points, these points are tripled. The trustee then decides how many of his 2000 points to send back to the trustor. The trustee can return any amount in steps of 10 points. The instructions make it very clear that no participant plays the trust game with his neighbor.

In the third part, each participant chooses the probability with which information about his pro-sociality in the previous part is revealed to his neighbors. Participants first receive feedback concerning the relative level of their donation or return decision as trustee. Everybody is randomly matched with two other participants.[2] In the treatments with donations, participants are informed whether their donation is higher, smaller, or neither higher nor smaller than the donation of the other two participants. Before their relative feedback, they guess the relative level of their donation. This belief question is not incentivized. Essentially the same is done in the treatment with the trust game, where people are ranked according to their return behavior.

Participants are next assigned the color green, yellow, or red. Colors reflect their relative pro-sociality in the treatment Altruism and Trust. In Altruism, a participant is assigned green if his donation is higher, red if his donation is smaller than the donation of the other two randomly assigned participants, and otherwise yellow. Colors in these treatments have a clear meaning: red is bad, green is good, and yellow is in between. In the control treatment Random, colors are assigned randomly, each color with equal probability. The instructions clarify the assignment and meaning of the colors. Understanding is ensured with the help of computerized control questions that have to be answered correctly by everybody before the third part of the experiment.

---

[2]Participants are never matched to their neighbors, and neighbors are never matched with the same other participants. Participants thus never learn something from their own relative feedback about the pro-sociality of their neighbor. This is made clear in the instructions.

Participants then determine the probability with which their own color is revealed to their neighbor. The revelation probability is chosen from 0.00 to 1.00 in steps of 0.05. Setting the probability to fifty percent is costless. Increasing or decreasing the revelation probability is possible at moderate costs. The cost function is symmetric around 50 percent and strictly convex. Adjusting the revelation probability for the first 5 percentage points costs 5 points. Revelation can be ensured or prevented for sure by spending 120 points. The full cost function can be found in the instructions in the appendix. Participants never directly observe the revelation probability set by their neighbor. This makes it less likely that people do not adjust the revelation probability, because they do not want to reveal that they care for their social image.

Participants finally learn the color of their neighbor with the revelation probability set by their neighbor. They then report how likable and pro-social they find their neighbor. Pro-sociality is defined as altruism in the treatments Altruism and Trust, and as trustworthiness in the treatment Trust. Participants are assured that these non-incentivized assessments are not revealed. Participants finally answer a questionnaire concerning their age, gender, field of study, social engagement, donation behavior, and experience with economic experiments.

For ethical reasons, participants receive written instructions describing the overall outline of the experiment with reference to all parts at the very beginning of the experiment. Everybody is thus informed that some information on their pro-social behavior in the first part might be revealed to others, but that this information has no instrumental value. Participants receive a detailed description before each corresponding part to prevent cognitive overload.

The experiment was programmed in z-tree by Fischbacher (2007) and conducted in the FLEX laboratory at the Goethe University of Frankfurt. Subjects were recruited via ORSEE by Greiner (2015). Between 6 to 20 subjects participated in 38 sessions, in total 590 subjects from a standard student subject pool. Total earnings equaled the sum of the earnings in the three parts of the experiment. The conversion rate was 1 eurocent per point. Subjects received in addition a show-up fee of 4 euros. Participants earned on average 11.69 euros for about 45 minutes. An English translation of the originally German instructions can be found in the appendix.

3 Theoretical Predictions

This section develops a simple theory to clarify thoughts. The experiment is designed to create an intuitive decision environment where it is very clear what participants should do if they want to manipulate their social image. Colors have a clear meaning in Altruism and Trust: those assigned green are on average more pro-social than those assigned yellow, who in turn are on average more pro-social than those assigned red. If people want to manipulate the pro-social impression they make on others, they should set a higher revelation probability if assigned green rather than yellow, and if they are assigned yellow rather than red. Colors have no meaning in Random. The revelation probability should then not condition on the assigned color. The following theoretical analysis makes these predictions more precise. It shows that the experimental design generates specific and clear predictions without requiring equilibrium arguments or strong assumptions on how participant believe their behavior is interpreted by other participants.

To keep the analysis simple, suppose participants have one of two types $\theta \in \{d, u\}$ that are called desirable and undesirable. These types are defined by their initial pro-social behavior $a(\theta)$ which is either the donation or the return behavior in the trust game. Desirable types are by definition more pro-social and thus $a(d) > a(u)$. Types are initially private information, where it is commonly known that all participants have the desirable type d with equal probability $\mu \in]0, 1[$. Suppose that all participants want to convince their neighbors that they have the desirable type, hence the denomination desirable and undesirable.

Participants might want to affect the beliefs of their neighbors by revealing their assigned color $c \in \{g, y, r\}$. Setting a specific revelation probability p can be costly. The cost function f is strictly convex and symmetric around 0.5 with $f(0.5)$ normalized to zero. Participants only spend resources to set the revelation probability p if they believe that revealing a particular color changes the pro-social impression they make on their neighbor. Let $\nu(c)$ be the second-order belief of participants revealing their assigned color $c \in \{g, y, r\}$. Belief $\nu(c)$ thus is the probability with which these participants believe their neighbors to believe them to have the desirable type. Let $\nu(n)$ be the second-order belief if participants do not reveal their color to their neighbors, where n stands for nothing revealed.

Consider a participant who is assigned color $c \in \{g, y, r\}$ and then sets revelation probability p. Define the expected utility of this participant as

$$p\,\nu(c)\,k + (1-p)\,\nu(n)\,k - f(p) \tag{1}$$

where $k > 0$ measures the strength of social image concerns. If participants maximize their expected utility and there is an interior solution, then the first order conditions

$$\big(\nu(c) - \nu(n)\big)\,k - f'\big(p^*(c)\big) = 0 \tag{2}$$

implicitly define optimal revelation probabilities $p^*(c)$ for each assigned color. Convexity of the cost function f implies that the optimal revelation probability is increasing in the difference in second-order beliefs $\nu(c) - \nu(n)$ and consequently increasing in $\nu(c)$ for any given $\nu(n)$. This is the main prediction of the model.

The design advantage of the experiment is that colors have a specific meaning which is determined by the experimental design and not open to interpretation or speculation. In the above simple theory, $\nu(g)$ in Altruism and Trust must be one, because only desirable types are ever assigned green, and $\nu(r)$ must be zero because only undesirable types are ever assigned red. The belief $\nu(y)$ lies strictly in between zero and one if the prior μ is strictly between zero and one. Then $\nu(g) > \nu(y) > \nu(r)$ holds which implies $p^*(g) > p^*(y) > p^*(r)$. In Random, colors are randomly assigned and have no informational content. This implies $\nu(g) = \nu(y) = \nu(r) = \mu$ and consequently $p^*(g) = p^*(y) = p^*(r)$.

Summarizing, if participants care for their social image, then they should condition their revelation probabilities on colors in Altruism and Trust. In particular, they should set a higher revelation probability if assigned green rather than yellow, and if assigned yellow rather than red. Participants should not condition their revelation probabilities on colors in Random, because colors in Random have no meaning.

Note that the theory only predicts a relative ordering of revelation probabilities. Without additional assumptions, it does not generate further predictions, because absolute revelation probabilities can depend on the second-order belief $\nu(n)$. This second-order belief describes how participants believe their neighbors to believe how other participants condition their revelation probability on assigned colors. For example, if people believe that their neighbors believe that those assigned green or yellow set a very high revelation probability, then not

observing the assigned color leads to the interpretation that the color must be red. Then not revealing any color is a very bad signal. What second-order beliefs participants hold is unclear. Therefore, the theory does not predict that participants should leave their revelation probability at 0.5 if assigned the neutral color yellow, because not revealing any color could be interpreted as a bad signal. Equally, the theory does not predict that participants should always set the cost-minimizing revelation probability of 0.5 in Random, because in this treatment the revelation of colors as such can have informational content.

4 Experimental Results

This section first studies whether participants condition revelation probabilities on assigned colors in the two treatments with meaningful colors. It then investigates how participants interpret the colors assigned to their neighbors. It finally explores treatment differences between Altruism and Random and studies the link between the revelation probabilities and colors after controlling for pro-social behavior, thereby addressing a potential endogeneity problem.

Participants should condition their revelation probability on meaningful colors if they want to manipulate the pro-social impression they make on others. In particular, they should set a higher revelation probability if assigned a color with a more favorable informational content. Figure 1 shows for each treatment the average revelation probabilities conditional on color. Focusing first on the treatments with meaningful colors, the average revelation probabilities are 57% for green, 50% for yellow, and 46% for red in Altruism. They are 64% for green, 56% for yellow, and 45% for red in Trust.

Two observations are immediate. First, participants condition their revelation probabilities on the assigned and meaningful colors in Altruism and Trust. Second, average revelation probabilities are close to 50%. Social signaling is not very pronounced in the experiment. Of all participants, 78% in Altruism and 66% in Trust do not spend any material resources to increase or decrease the revelation probability. Only a minority of participants engages in strategic information revelation. However, these participants seem to care for the social impression they make on others, just as suggested by the theory.
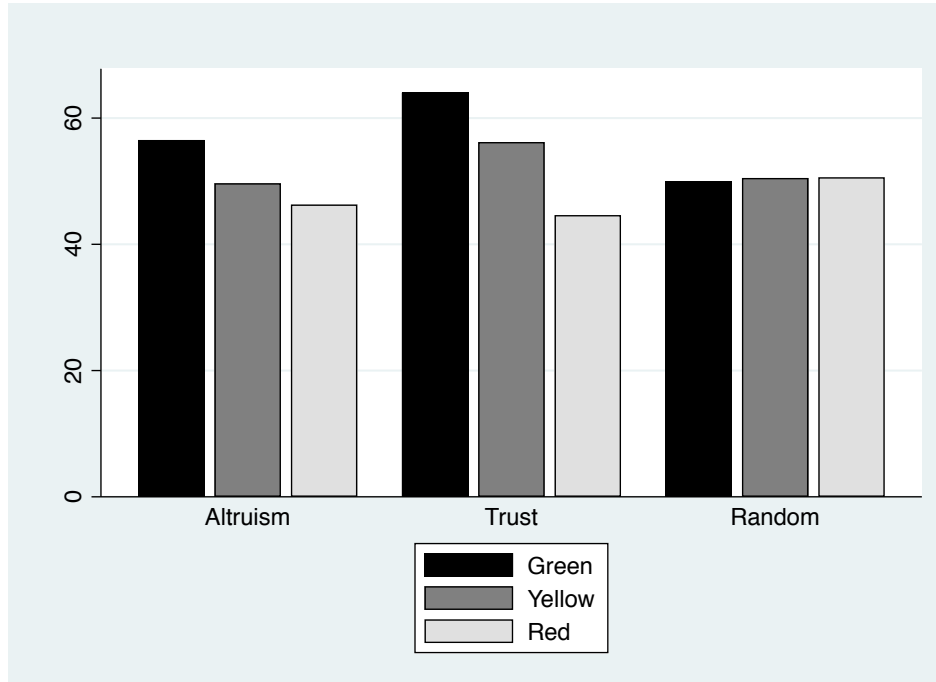
Figure 1: Average Revelation Probabilities Conditional on Colors

Notes: The figure shows average revelation probabilities conditional on the assigned color for all treatments. Colors reveal information on relative past donations in Altruism and past trustworthiness in Trust. Colors in Random are randomly assigned and thus have no informational content.

Regression analysis confirms these observations. Figure 2 reports the results from regression analysis. The baseline regressions simply regress the revelation probability on dummies for being assigned green or red, with no further control variables. The omitted reference category is being assigned the color yellow. The estimated coefficients of the baseline regressions are plotted as black dots, with 95% confidence intervals clustered on pairs. Clustering is necessary because participants interact in pairs before making their pro-social decisions and before setting their revelation probabilities, so that observations within pairs cannot be considered independent. All p-values reported in the paper are rounded to two digits and based on two-sided tests.
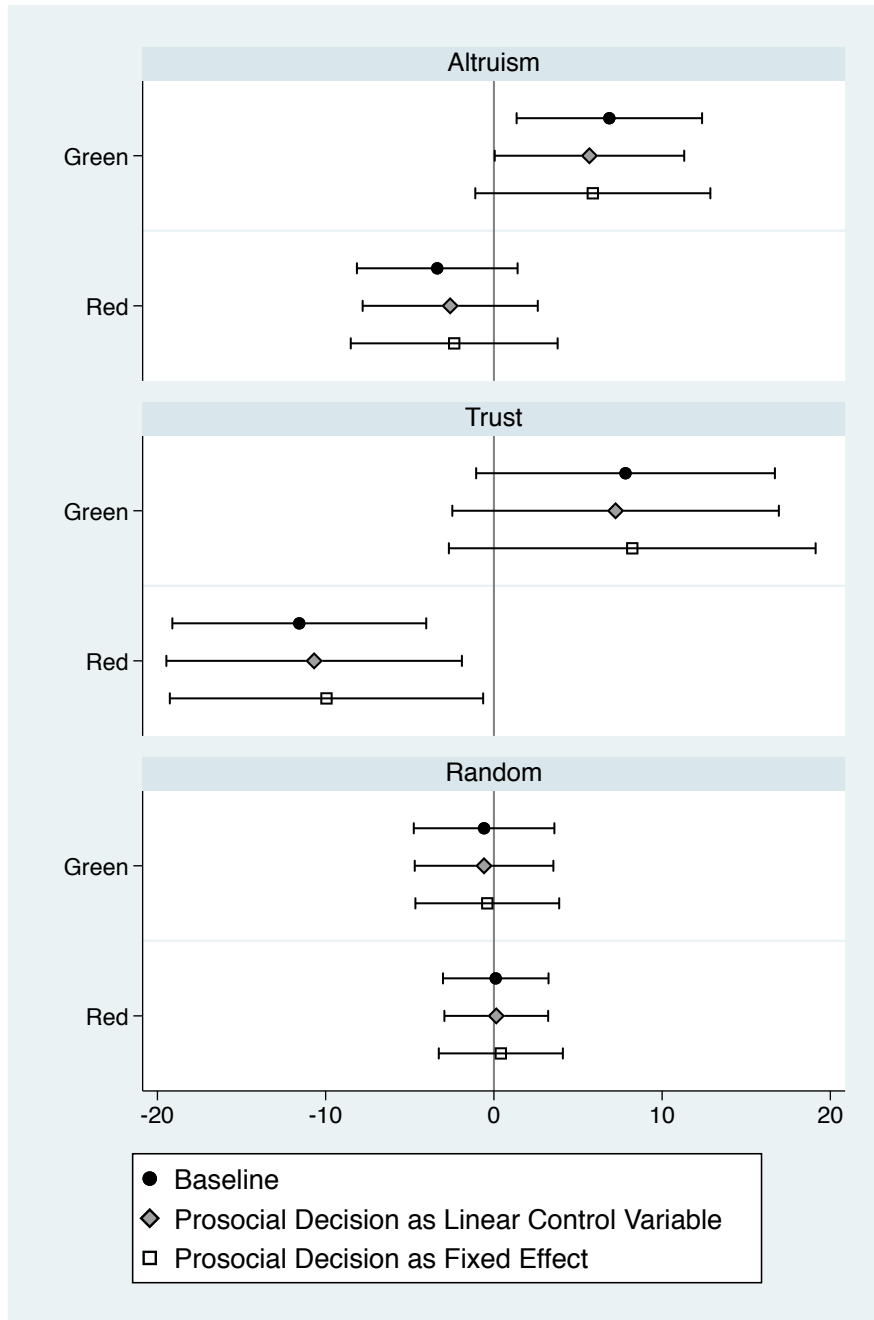
Figure 2: Regression Results

Notes: The figure plots the estimated coefficients of linear regressions with revelation probability as dependent variable. Green and Red refer to dummy variables indicating the assigned color, the omitted reference category is the color yellow. The three panels refer to the treatments Altruism, Trust and Random. The reported point estimates come from regressions that do not control for the initial pro-social decision (points), that include the initial pro-social decision as a linear control variable (diamonds), and that include pro-social decision fixed effects (squares). Pro-social decisions are donations in Altruism and Random, and trustworthiness in Trust. The bars represent 0.95 confidence intervals, in all regressions standard errors are clustered on pairs. The numbers of observations are 220 in Altruism, 156 in Trust, and 214 in Random.

13

The baseline regressions indicate that participants in Altruism increase their revelation probability if assigned green rather than yellow, but not if assigned red rather than yellow (p-values of 0.02 and 0.17). A Wald test confirms that participants set a higher revelation probability if assigned green rather than red (p-value of 0.00). Participants in Trust increase their revelation probability if assigned green rather than yellow, but this effect is only weakly significant (p-value of 0.08). Participants decrease the revelation probability if assigned red rather than yellow (p-values of 0.00). A Wald test again confirms that participants set a higher revelation probability if assigned green rather than red (p-value of 0.00).

Non-parametric robustness tests generate the same conclusion. Because observations within pairs cannot be considered independent, the non-parametric analysis follows a bootstrapping procedure. Within every pair, one randomly selected participant enters the statistical test. This procedure is repeated 1000 times with new random draws. Reported are the average p-values and the rate at which the null hypothesis is rejected at the 10% significance level. Any bootstrapped non-parametric test is said to reject the respective null hypothesis if the average p-value is less than 10%. Using this bootstrapping procedure, Kruskal-Wallace tests reject that participants do not condition their revelation probabilities on colors in Altruism and Trust (average p-values of 0.04 and 0.03 with rejection rates of 0.89 and 0.92).

The empirical results so far suggest that participants strategically manipulate the pro-social impression they make on relative strangers. But manipulating the revelation probabilities conditional on color only affects social image if colors are interpreted in the right way. Concerning meaning and interpretation, colors in Altruism and Trust have per design a specified meaning – those with green behaved more pro-socially than two randomly chosen other participants, and so on for the other two colors.

The data show that the experimental design worked as intended. Bootstrapped Spearman rank correlation tests show that colors are correlated with pro-social behavior in Altruism and Trust (average p-values of 0.00 with rejection rates of 1.00). Regression analysis confirms that participants assess their neighbors as more pro-social if the revealed color is more favorable (p-values of 0.00 with standard errors clustered on pairs). Colors have the intended informational content and are interpreted accordingly in Altruism and Trust. The signaling story finally requires that participants form the right beliefs about how colors are interpreted by their neighbors. Beliefs are determined by design and understanding of the

rules in the experiment are ensured via control questions that had to be correctly answered by all participants. It therefore is highly likely that the common interpretation of colors is common knowledge.

The remaining section investigates the following potential endogeneity problem. Pro-social behavior and thereby color assignment are endogenous in Altruism and Trust. It is thus possible that revelation probabilities and colors are correlated, not because participants want to impress their neighbors, but because pro-social participants have unobservable characteristics that affect both their pro-social behavior and their choice of revelation probability. For example, pro-social participants could have a preference for transparency. Such participants would then set high revelation probabilities and at the same time act pro-socially, which makes it incidentally more likely to be assigned a favorable color. The unobservable characteristic then creates correlation between colors and revelation probabilities, although people do not care for their social image.

The first way to to address the above potential endogeneity concern is comparing behavior in Altruism and Random. The treatments Random and Altruism are identical, with the only difference that colors are randomly assigned and thus have no informational content in Random. If some unobservable characteristics drive both donations and revelation probabilities, a similar correlation between colors and revelation probabilities should be found in both Altruism and Random. Bootstrapped Spearman rank correlation tests show that donations and revelation probabilities are correlated in Altruism (average p-value of 0.09 with rejection rate of 0.74) but not in Random (average p-value of 0.44 with rejection rate of 0.11). This already indicates that behavior in Altruism is driven by social image concerns.

More importantly, the theory predicts that participants who care for their social image should condition their revelation probabilities on colors if and only if these colors have informational content. In Random, colors have no informational content. Figure 1 shows that the average revelation probabilities in Random are 50% for green, 51% for yellow, and 51% for red. The revelation probabilities are essentially the same and very close to 50% for all three colors. The baseline regression reported in the bottom panel of Figure 2 finds that participants in Random do not condition their revelation probability on colors (p-values of at least 0.74). A bootstrapped Kruskal-Wallace confirms this result (average p-value of 0.55

15

with rejection a rate of 0.05). Finally, 82% of participants in Random leave their revelation probability at the cost-minimizing 50%.

Participants therefore do not spend material resources to affect the revelation probability of colors if these colors have no meaning. Participants also understand that colors have no meaning in Random. Bootstrapped Spearman rank correlation tests show that colors in Random are not correlated with pro-social behavior (average p-value of 0.58 with rejection rate of 0.03) and therefore have no informational content. Regression analysis shows that revealed colors do not affect how participants assess the pro-sociality of their neighbors in Random (p-values weakly larger than 0.27).

Using treatment comparisons as identification strategy assumes that due to random treatment assignment, participants do not systematically differ in observable and unobservable characteristics across the treatments. Concerning exogenous background variables, there are almost no significant differences between Altruism and Random (p-values weakly larger than 0.35). The only exception is that there are more participants studying economics in Altruism than in Random (p-value of 0.00). The reason is that some sessions of Random were run during the semester break – not a great idea ex-post. Because the laboratory is situated close to the Faculty of Economics and Business, sessions in the semester break attract relatively fewer participants studying economics. There are no differences in any background variables between Random and Altruism when considering only sessions run during the semester (p-values weakly larger than 0.49). Kruskal-Wallace tests reveal that participants in Random from these comparable sessions during the semester do not condition their revelation probabilities on their colors (average p-value of 0.60 with rejection rate of 0.03). Regression analysis not further reported here leads to the same conclusion.

Concerning the endogenous donation decision, it actually turns out that average donations are 353 in Altruism and thus substantially lower than the 419 in Random. This difference in donations must be coincidental because both treatments are identical – they use the same instructions and procedures – until after the donation. However, regression analysis reveals that the difference in donations is significant (p-value of 0.05). Ranksum tests yield a more ambiguous result (average p-value of 0.16 with rejection rate of 0.55) which already suggests that the difference could be driven by few observations. Inspection of the data shows that the two sessions with the largest average donations of 543 and 588 both happen

to fall into Random, one during the semester, and one during the semester break. These two sessions are outliers, the next highest average donation is substantially lower at 464. The null hypothesis of equal donations in Altruism and Random can no longer be rejected when excluding the two outlier sessions, both by ranksum tests (average p-value of 0.32 with rejection rate of 0.20) and regression analysis (p-value of 0.18). Kruskall-Wallace tests excluding the two outlier sessions find that participants do not condition their revelation probabilities on colors in Random (average p-value of 0.59 with rejection rate of 0.02). Regression analysis not further reported here leads to the same conclusion.

Comparing behavior in Altruism and Random therefore suggests that participants really care for the pro-social impression that they make on their neighbors. For those not yet convinced, an alternative approach to address the potential endogeneity problem employs regression analysis. This regression analysis uses the random composition of the assigned reference groups for identification. Given any pro-social behavior, the relative ranking of participants is determined by the randomly assigned reference group. The section next explores whether participants condition their revelation probabilities on colors once controlling for pro-social behavior.

Figure 2 reports the results from regressions that include pro-social decisions as linear control variable, where the estimated coefficients are plotted as grey diamonds. Participants in Altruism increase their revelation probabilities if assigned green rather than yellow (p-value of 0.05). A Wald test reveals that participants set a high revelation probability if assigned green rather than red (p-value of 0.01). Participants in Trust decrease their revelation probability if assigned red rather than yellow (p-value of 0.02). A Wald test shows that participants set a lower revelation probability is assigned red rather than green (p-value of 0.01). Conclusions are the same from regressions with decision fixed effects, where the estimated coefficients are reported as hollow squares. Controlling for pro-social decisions does not affect the regression analysis very much, except for slightly inflating standard errors. Closer inspection actually shows that there is no clear and systematic link between pro-social behavior and revelation probabilities, once controlling for the assigned color in Altruism in Trust. Regression analysis therefore corroborates the results from the initial Spearman rank correlation tests, which find no significant correlation between pro-social behavior and revelation probabilities in Random. It seems unlikely that some unobservable characteristics drive both pro-social behavior and color revelation.

Finally, regression analysis controlling for pro-social decisions yields that participants do not condition their probabilities on colors in Random (p-values weakly larger than 0.78). This regression analysis combines treatment comparisons and controlling for pro-social decisions to address the potential endogeneity problem.

5   Conclusion

The present experiment shows that people care what you think, and therefore confirms conventional wisdom and the existing experimental literature on social image concerns. Going beyond the existing empirical evidence, the experiment indicates that people very strategically and thus consciously manage their social image. This comforting insight lends further credibility to theoretical social signaling models, which essentially assume that people correctly anticipate how their behavior affects their social reputation, see for example Bénabou & Tirole (2006) and Ellingsen & Johannesson (2008).

Social image concerns can strongly affect the efficiency of organizations, see for example the use of symbolic awards in Kosfeld & Neckermann (2011). The present research design could be easily adapted to study whether social image concerns influence all kinds of behavior, for example risk-seeking, perseverance, or norm compliance. Such knowledge can have important managerial implications. Furthermore, the present experimental paradigm can provide new information on what exactly people want to signal. The current data tentatively suggest that people like to reveal high altruism, but do not care so much for hiding meanness, and that they want to hide betrayal, rather than show off trustworthiness. The standard paradigm – varying observability – cannot detect such subtle mechanisms of social pressure. The reason is that people acting under scrutiny can both reveal positive and hide negative characteristics only by changing behavior in the same direction, for example by making a higher donation. But as Bursztyn & Jensen (2017) argue, future insights into the precise mechanisms of social image concerns and social pressure could provide important information for the design of effective public policies, organizational structures, and incentive systems.

References

Andreoni, J. & Bernheim, B. D. (2009). Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects, *Econometrica* **77**(5): 1607–1636.

Ariely, D., Bracha, A. & Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially, *American Economic Review* **99**(1): 544–555.

Austin-Smith, D. & Fryer, Jr., R. G. (2005). An economic analysis of "acting white", *Quarterly Journal of Economics* **120**(2): 551–583.

Bateson, M., Nettle, D. & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting, *Biology Letters* **2**: 412–414.

Bénabou, R. & Tirole, J. (2006). Incentives and prosocial behavior, *American Economic Review* **96**(5): 1652–1678.

Burnham, T. C. & Hare, B. (2007). Engeneering human cooperation - does involuntary neural activation increase public goods contributions?, *Human Nature* **18**: 88–108.

Bursztyn, L. & Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding and shaping social pressure, *Annual Review of Economics* **9**: 131–153.

Dana, J., Weber, R. A. & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness, *Economic Theory* **33**(1): 67–80.

Dickinson, D. & Villeval, M.-C. (2008). Does monitoring decrease work effort? the complementarity between agency and crowding-out theories, *Games and Economic Behavior* **63**: 56–76.

Diener, E. & Wallbom, M. (1976). Effects of self-awareness on antinormative behavior, *Journal of Research in Personality* **10**: 107–111.

Eil, D. & Rao, J. M. (2011). The good news - bad news effect: Asymmetric processing of objective information about yourself, *American Economic Journal: Microeconomics* **3**: 114–138.

Ellingsen, T. & Johannesson, M. (2008). Pride and prejudice: The human side of incentive theory, *American Economic Review* **98**(3): 990–1008.

Falk, A. (2017). Facing yourself - a note on self-image, mimeo, University of Bonn pp. 1–18.

Falk, A. & Ichino, A. (2006). Clean evidence on peer effects, *Journal of Labor Economics* **24**(1): 39–57.

Fehr, E. & Schneider, F. (2010). Eyes are on us, but nobody cares: Are eye cues relevant for strong reciprocity?, *Proceedings of the Royal Society Biology* **277**(1686): 1315–1323.

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

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

Haley, K. J. & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game, *Evolution and Human Behavior* **26**: 245–256.

Kosfeld, M. & Neckermann, S. (2011). Getting more work for nothing? symbolic awards and worker performance, *American Economic Journal: Microeconomics* **3**(2): 86–99.

Mas, A. & Moretti, E. (2009). Peers at work, *American Economic Review* **99**(1): 112–145.

Mellström, C. & Johannesson, M. (2008). Crowding out in blood donation: Was titmuss right?, *Journal of the European Economic Association* **6**(4): 845–863.

Soetevent, A. R. (2005). Anonymity in giving in a natural context - a field experiment in 30 churches, *Journal of Public Economics* **89**: 2301–2323.

van der Weele, J., Kulisa, J., Kosfeld, M. & Friebel, G. (2014). Resisting moral wiggle room: How robust is reciprocal behavior?, *American Economic Journal: Microeconomics* **6**(3): 256–264.

Appendix: Instructions

General Remarks

Please find attached the instructions for the experiment. The parts that were specific for the different treatments are in brackets and in different color. The black parts are the same in all treatments. Subjects filled out a questionnaire at the end of the experiment. The questionnaire is not included. Subjects had to answer computerized answer control questions before the experiment could start. The control questions are included at the end. The remarks and comments were of course not included in the instructions.

Instructions

Many thanks for your participation in this experiment. The amount of money that you can earn depends on your own decisions and potentially also on the decisions of the other participants. During the experiment, you can earn points. 1 point corresponds to 1 Eurocent (100 points thus correspond to 1 Euro).

We will pay out the total amount of money you earned in cash at the end of the experiment. We guarantee that your earnings remain confidential. Consequently, we will not disclose any information on your earnings to other participants of the experiment.

Below you find the instructions of the experiment which you should read. Please do not communicate with any other participant during the experiment, except if it is allowed and requested according to the instructions. Please raise your hand when you have a question. One of the experimenters will come to your table to confidentially answer your question.

The Experiment

All participants have the same tasks in the experiment. For simplicity, we will refer to your decisions in the following instructions. Please keep in mind that all participants are in the same situation as you and thus face the same decisions as you.

[Altruism and Random] You interact with your neighbor in this experiment. The experiment consists of three parts. In the first part, at first you fill out a questionnaire. Afterwards, you talk with your neighbor. Thereafter, you fill out a questionnaire concerning your neighbor. In the second part, you decide on a donation for the "German children's cancer charity". In the third part, you get to know how much you donated compared to the other participants after a brief self-evaluation. Then you can decide on a probability with which your neighbor potentially gets some information on your relative donation. After you potentially have gotten some information on your neighbor's relative donation, you are asked to assess the altruism of your neighbor. Additionally, we want to know how congenial your partner is. You can earn points in every part of the experiment. Your total earnings are the sum of your earnings in each of the three parts plus an additional 400 points for participating in this experiment.

Part 1

In this experiment, you interact with your neighbor. At first, please fill out a questionnaire on your preferences. We ask you questions on colors, soccer teams, tress, days, and musical instruments. You cannot earn points by filling out the questionnaire.

After having completed the questionnaire, you have 5 minutes time to discuss the questionnaire with your neighbor. It is absolutely prohibited to communicate with anybody after the discussion until the end of the experiment. In particular, you are not allowed to talk with your neighbor until we have informed you that the experiment is over.

After having talked with your neighbor, please fill out the same questionnaire on the preferences of your neighbor at the PC. You can now earn points by correctly filling out the questionnaire: you get 20 points for each answer that corresponds to your neighbor's original answer.

Part 2

[Altruism and Random] In the second part of the experiment, you can donate money to the "German children's cancer charity." Below you find information from the website of the "German children's cancer charity".

The "German children's cancer charity foundation" was founded 1996 by the "German children's cancer charity." It helps affected persons with words and deeds and is heavily involved in fighting cancer during infancy. The "German cancer charity" and "German children's cancer charity" fund almost all current studies on therapies for children. The success of continuously-improving therapy concepts is impressive: Whereas, a leukemia diagnoses for a child almost certainly ended with death three decades ago, nowadays almost 80% of the young patients survive the disease. The "German cancer charity" funds its activities exclusively from donations and does not receive any public funds. (Source: Website of the "German children's cancer charity)

You receive 1000 points for your donation decision from us. You can decide how many of those points you want to donate to the "German children's cancer charity." The donation can be between 0 and 1000 points, with 5 points increments. As in the rest of the experiment, one point corresponds to one Eurocent.

Please be keep in mind that the donation decision is real! We will forward the monetary equivalent of your donation to the "German children's cancer charity" after the experiment. We will pay out the rest of the 1000 points, which you did not donate, to you at the end of the experiment. You can contact us if you want to receive additional information on the donations.

 [Trust] In the second part of the experiment, you play with a randomly chosen participant the so-called trust game. **We ensure that no participant plays with his or her neighbor**. Below we describe the trust game.

There are two player roles in this game: player A and player B. Both players start with an endowment of 500 points. Player A has to decide first whether he sends all or not any points to player B. If player A does not send any points to player B the game ends. But if player A sends all his points to player B, the number of points is tripled. Before player B makes a decision, player A then has no points, whereas player B has, together with his endowment, $500 + 3*500 = 2000$ points. Thereafter, player B has to decide how many of those 2000 points he wants to send back to player A. The number of points can be between 0 and 2000 points, with increments of 10 points. The number of points sent back to player A is not tripled and player A receives only one point for every point sent. Player A thus receives the points sent to him, whereas player B gets 2000 minus the number of points he sent to player A. Afterwards, the game ends. As in the rest of the experiment, one point corresponds to one Eurocent. The following examples illustrate the rules of the game.

Part 3

[Altruism] In this part of experiment, at first you will get information on the relative amount of your donation of the second part of the experiment. For this, we compare your donation with the donations of two randomly chosen other participants. You get to know whether you donated more than the two other participants, or whether you donated less than both other participants, or whether you donated neither more nor less than both other participants. We ensure that you are not compared with your neighbor. Furthermore, we will compare you and your neighbor with different participants. Consequently, you cannot learn anything on the relative amount of your neighbor's donation from the relative amount of your own donation. Similarly, your neighbor is not able to infer something about your relative donation amount based on his feedback.

We will ask you how high you assess your relative donation amount. You cannot earn points by answering this question. Despite this, we still ask you to truthfully answer the question.

Upon having received and confirmed the relative feedback, you can influence the information which you can potentially send to your neighbor about your donation. For this, we assign you one of the colors red, yellow, or green. Then you can decide on the probability with which your neighbor gets to know your color.

Your color is determined based on relative donation amount to the "German children's cancer charity." If you donated more than both - before mentioned – other comparison participants, you will receive the color green. If you donated less, the color red will be assigned to you. In all other cases – if you donate neither more nor less than both comparison participants, you will receive the color yellow. **We will assign a color to your neighbor the same way**. The following table summarizes the color assignment process:

Donation to "German children's cancer charity"	Assigned color
Higher than the donation of both your comparison participants	Green
Neither more nor less than the donation of both your comparison participants	Yellow
Less than the donation of both your comparison participants	Red

As already indicated above, we ensure that not any participant is compared with his neighbor. Consequently, your color and the color of your neighbor always are independent.

[Random] In this part of experiment, at first you will get information on the relative amount of your donation of the second part of the experiment. For this, we compare your donation with the donations of two randomly chosen other participants. You get to know whether you donated more than the two other participants, or whether you donated less than both other participants, or whether you donated neither more nor less than both other participants. We ensure that you are not compared with your neighbor. Furthermore, we will compare you and your neighbor with different participants. Consequently, you cannot learn anything on the relative amount of your neighbor's donation from the relative amount of your own donation. Similarly, your neighbor is not able to infer something about your relative donation amount based on his feedback.

We will ask you how high you assess your relative donation amount. You cannot earn points by answering this question. Despite this, we still ask you to truthfully answer the question.

Upon having received and confirmed the relative feedback, you can influence the information which you can potentially send to your neighbor about your donation. For this, we assign you one of the colors red, yellow, or green. Then you can decide on the probability with which your neighbor gets to know your color.

The color has no meaning in this experiment and is assigned completely randomly.
Every color has the same probability. Your color thus contains no information on your donation to the "German children's cancer charity". Similarly, your neighbor's color does not tell you anything about the relative amount of his donation.

Relative amount of the donation	Assigned color
Higher than the donation of both your comparison participants	completely random
Neither more nor less than the donation of both your comparison participants	completely random
Less than the donation of both your comparison participants	completely random

[Trust] In this part of experiment, at first you will get information on your relative trustworthiness of the second part of the experiment. For this, we compare your behavior as player B with the behavior of two randomly chosen other participants. You get to know whether you send back more to player A than the two other participants, or whether you sent back less than both other participants, or whether you sent back neither more nor less than both other participants. We ensure that you are not compared with your neighbor. Furthermore, we will compare you and your neighbor with different participants. Consequently, you cannot learn anything on the relative trustworthiness of your neighbor from your own relative trustworthiness. Similarly, your neighbor is not able to infer something about your relative trustworthiness amount based on his feedback.

We will ask you how high you assess your relative trustworthiness amount. You cannot earn points by answering this question. Despite this, we still ask you to truthfully answer the question.

Upon having received and confirmed the relative feedback, you can influence the information which you can potentially send to your neighbor about your trustworthiness. For this, we assign you one of the colors red, yellow, or green. Then you can decide on the probability with which your neighbor gets to know your color.

Your color is determined based on your relative trustworthiness. If you, as player B, wanted to send back more to player A than both of your – afore mentioned – comparison participants, you will get the color green. If you wanted to send back less, you will be assigned the color red. In all other cases, if you wanted to send back neither more nor less, you will get the color yellow. **We assign your neighbor's color the same way**. The following table summarizes the color assignment process:

First, we will tell you your [Random] randomly determined color. After having confirmed this, you can choose the probability with which your neighbor gets to know your own color. Your probability of revealing your color must be between 0% and 100%, with 5% increments.

If you set the probability exactly to 50%, you incur no costs. Apart from that, setting a specific probability costs points. The more your probability of revealing your color deviates from 50%, the higher the costs. The following table gives you the exact costs in terms of points. As in the rest of the experiment, one point corresponds to one Eurocent.

Probability	Costs (points)
0%	120
5%	90
10%	70
15%	50
20%	40
25%	30
30%	20
35%	15
40%	10
45%	5
50%	0
55%	5
60%	10
65%	15
70%	20
75%	30
80%	40
85%	50
90%	70
95%	90
100%	120

Example 1: Suppose you want to set the probability to 40%. Then you will incur costs of 10 points.

Example 2: Suppose you want to set the probability to 65%. Then you will incur costs of 15 points.

Example 3: Suppose you want to set the probability to 50%. Then you will not incur any costs.

Your neighbor chooses the probability with which you will receive his color the same way. After all decisions have been made, your color will be revealed to your neighbor according to your chosen probability. You either get to know the color of your neighbor or you don't get to know it. You do not receive any additional information. In particular, you never get to know the probability your neighbor chose. Thus, you cannot see whether your neighbor chose a probability of 35% or 60%. Similarly, your neighbor is – of course – not able to see your chosen color. [Altruism] Furthermore, the participants never receive any information on the absolute amount of donation of their neighbor. [Random] Furthermore, the participants never receive any information on the absolute amount of donation of their neighbor. [Trust] Furthermore, the participants never receive any information on the absolute number of points that their neighbor, as player B, wanted to send back to player A.

After you potentially got to know your neighbor's color, you have to evaluate the [Altruism] degree of altruism [Random] degree of altruism [Trust] trustworthiness of your neighbor. Additionally, you should indicate, how congenial you find your neighbor. You cannot earn any points with those assessments. Despite this, we ask you to truthfully answer the questions. Furthermore, we keep those assessments strictly confidential. In particular, your neighbor will never get to know your assessment.

Control Questions First Screen

Before the third part of the experiment, you have to correctly answer some control questions. Afterwards, the last part of the experiment begins.

What can you say about a participant who has been assigned the color green?
What can you say about a participant who has been assigned the color yellow?
What can you say about a participant who has been assigned the color red?

[Altruism] The options refer to donation behavior.
[Trust] The options refer to donation behavior.
[Random] The options refer to donation behavior.

[Comments] There are four radio buttons as options. They refer to donations or returns in the trust game that are higher than the average, about the average, and lower than the average. The fourth option is that one cannot say much about the behavior of the participant. The correct answers follow from the treatment, where the fourth option is the correct answer for all three colors in the treatment Random.

Will your neighbor ever learn your chosen revelation probability?
What are your costs if you chose a revelation probability of 45%?
What are your costs if you chose a revelation probability of 70%?
What are your costs if you chose a revelation probability of 50%?

[Comment] Options for the first question are Yes and No. The right answer is No in all treatments. For the remaining three questions, the right costs had to be entered. The experiment only started once all questions had been answered correctly.