Journal of the Economic Science Association (2020) 6:43–56
https://doi.org/10.1007/s40881-020-00085-2

ORIGINAL PAPER

I care what you think: social image concerns and the strategic revelation of past pro-social behavior

Ferdinand A. von Siemens[1,2]

Received: 7 August 2019 / Revised: 27 January 2020 / Accepted: 24 March 2020 /
Published online: 19 April 2020

Abstract

This article studies whether people want to control what information on their own past pro-social behavior is revealed to others. Participants are assigned a color that depends on their past pro-social behavior. They can spend money to manipulate the probability with which their color is revealed to another participant. The data show that participants are more likely to reveal colors with more favorable informational content. This pattern is not found in a control treatment in which colors are randomly assigned, thus revealing nothing about past pro-social behavior. Regression analysis confirms these findings, also when controlling for past pro-social behavior. These results complement the existing empirical evidence, confirming that people strategically and, therefore, consciously manipulate their social image.

Keywords Social signaling · Altruism · Trustworthiness

JEL Classification C91 · D03 · D83 · D82

"I was told when I get older all my fears would shrink. But now I'm insecure, and I care what people think. My name's Blurryface, and I care what you think."
Stressed Out, Twenty One Pilots.

Very helpful comments from participants in the Grueneburgseminar and excellent research assistance from Victor Klockmann and Alicia von Schenk are gratefully acknowledged.

✉ Ferdinand A. von Siemens
vonsiemens@econ.uni-frankfurt.de

[1] Faculty of Economics and Business Administration, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany

[2] CESifo, Poschingerstrasse 5, 82679 Munich, Germany

1 Introduction

Field and laboratory experiments suggest that people strategically manipulate and, therefore, consciously care about their social image. Social image concerns seem to influence a wide range of behaviors, such as charitable giving, workplace conduct, voting, consumption choices, financial decisions, and investments in education, see Soetevent (2005), Falk and Ichino (2006), Andreoni and Bernheim (2009), Ariely et al. (2009), Mas and Moretti (2009), and the survey by Bursztyn and Jensen (2017). However, almost all existing studies on social image concerns use the same empirical identification strategy: they argue that people care about their social image, because they change their behavior under the scrutiny of a human audience.

Bursztyn and Jensen (2017) argue that changing observability might change behavior through channels other than social image concerns. These channels could be privacy concerns, social learning, externalities, and concurrent changes in the decision environment. Observability might also influence behavior by making social norms more salient, see Mazar et al. (2008), or by increasing self-awareness, see Diener and Wallbom (1976) and Falk (2017). Rather than looking at how exogenously imposed observability changes behavior, the present paper investigates whether people themselves want to vary observability of their own past actions. The experiment thereby tests whether people deliberately and consciously control the flow of information to others and thereby their pro-social reputation.

In the treatment "Altruism," participants interact face-to-face with their neighbor. They then make a private charitable donation. Participants are next assigned a color: green if they donated more than two randomly determined other participants, red if they donated less, and otherwise yellow. Participants then determine at some cost the probability with which they reveal their color to their neighbor. Finally, participants learn the color of their neighbor with the probability determined by the latter.

Colors in Altruism convey by design certain information: green indicates more pro-social past behavior than yellow, which in turn suggests more pro-social behavior than red. If participants consciously care about their social image, they should set a high revelation probability if assigned green, an intermediate probability if assigned yellow, and a low probability if assigned red.

The data show that participants systematically condition their revelation probabilities on colors in Altruism. Average revelation probabilities are 57% for green, 50% for yellow, and 46% for red. However, this is not enough to conclude that people consciously care about 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 endogenous, because it depends on past pro-social behavior. Due to unobservable characteristics, participants who behaved pro-socially might also set a high revelation probability, for example, because pro-social participants also value transparency. Pro-social participants are typically assigned favorable colors. Colors and revelation probabilities are then correlated, although participants do not care about their social image.

The paper pursues two empirical approaches to deal with this potential endogeneity problem. First, it compares behavior in Altruism and a control treatment "Random". This control treatment Random is identical to Altruism, with the only difference that colors in Random are randomly assigned and thus reveal no information on past pro-social behavior. If participants strategically want to manipulate their social image, they should condition revelation probabilities on colors only in Altruism but not in Random. Average revelation probabilities in Random are essentially 50% for all three colors.

Second, regression analysis can tackle the potential endogeneity problem by controlling for the pro-social behavior of participants. Such regression analysis is possible, because assigned colors do not only depend on the donations but also the pro-social behavior of the randomly composed reference groups. Participants with the same donation are consequently sometimes assigned different colors. See Eil and Rao (2011) for a similar approach studying the effect of positive and negative feedback on self-image. Statistical analysis confirms that participants in Altruism condition revelation probabilities on colors, also when controlling for pro-social behavior.

The present experimental paradigm can be easily adapted to study social image concerns in almost any domain. A third treatment "Trustworthiness" illustrates the portability of the design. Participants are assigned colors conditional on their relative trustworthiness in a trust game. If participants care about their social image concerning their trustworthiness, they should condition revelation probabilities on colors. Average revelation probabilities are 64% for green, 56% for yellow, and 45% for red. Participants consciously want to convince others not only of their altruism but also of their trustworthiness.

The present paper complements a small number of papers that look at how people want to manipulate the observability of their past behavior. DellaVigna et al. (2017) study whether people want to talk to others about past voting behavior. Eriksson et al. (2017) find that participants in a laboratory experiment pay to avoid public exposure of the least performer in their group. Bursztyn et al. (2018) show that people pay more for credit cards which can signal high income and that demand for these credit cards drops if they become available to lower income groups. Holm and Samahita (2018) document that people are willing to incur costs to manipulate whether information on their contribution to a public good is published on the web.

These papers conclude that people care for their social image, because they spend material resources to control whether past behavior is revealed to—or discussed with—others. The results of the present paper point in the same direction. Apart from confirming the importance of social image concerns, the present paper makes two contributions to the existing literature. The first contribution is more methodological: only the present paper generates exogenous variation in the information people might want to reveal or hide. As argued before, if unobservable characteristics drive both pro-social behavior and subsequent information revelation, the observed correlation might be unrelated to social image concerns. For example, people who contributed a lot to the public good in Holm and Samahita (2018) might also have a preference for being transparent. In the present paper, even very pro-social participants are sometimes assigned an unfavorable color. Controlling for past pro-social

behavior—and comparison to the treatment Random—allows eliminating any potential endogeneity problem by design.

Concerning the second contribution, the present experimental paradigm provides 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 low altruism; and that they want to hide betrayal rather than show off trustworthiness. Just 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 and Jensen (2017) argue, future insights into the precise mechanisms of social image concerns and social pressure could provide valuable information for the design of effective public policies, organizational structures, and incentive systems.

2 Experimental design and procedures

The experiment consists of three parts. The first part increases social proximity between participants to strengthen social image concerns; see Dickinson and Villeval (2008) for a similar approach. Participants first fill out a questionnaire asking for their favorite color, football club, tree, and other items with no apparent connection to pro-sociality. Participants cannot earn anything with the first questionnaire. They then get 5 min to discuss the questionnaire face-to-face with their neighbor and afterward turn back to their cubicles. From that moment onwards, participants are no longer allowed to communicate directly. Participants then fill out each other's questionnaire. They receive 20 points for each answer equal to the initial response of their neighbor.

In the second part, participants face a pro-social decision. There are three treatments. In "Altruism" and "Random," participants divide 1000 points between themselves and a well known German charity (Deutsche Kinderkrebshilfe). Instructions stress that the charity is financed exclusively by donations, and they feature a large red slogan emphasizing the importance of contributions. In the treatment "Trustworthiness," 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. If the trustor sends no points, the game ends, and the trustor and trustee keep their endowments. If the trustor sends all his points, these are tripled. The trustee then decides how many points to send back. The instructions make it clear that nobody plays the trust game with his neighbor.

In the third part, participants choose with which probability to reveal information about their pro-sociality to their neighbor. Participants are first randomly matched with two other participants.[1] In the treatments with donations, participants learn

[1] Participants are never matched to their neighbors, and neighbors are never matched with the same other participants. Participants thus never learn anything from their relative feedback about the pro-sociality of their neighbor.

whether their donation is higher, smaller, or in between the donations of the other two participants. Before their feedback, they guess their relative pro-sociality. This belief question is not incentivized. Procedures are the same in the treatment Trustworthiness; people there are ranked according to their return behavior.

Participants are then assigned the color green, yellow, or red. Colors reflect relative pro-sociality in Altruism and Trustworthiness. In Altruism, participants get green if their donation is higher, red if their donation is smaller than the donation of the other two randomly assigned participants, and otherwise yellow. Colors have a definite meaning: red is bad news, green is good news, and yellow is in between. In Random, colors are assigned randomly, each with equal probability. The instructions and computerized control questions clarify the assignment and the meaning of the colors.

Participants finally determine the probability with which to reveal their color to their neighbor. The revelation probability is chosen from 0.00 to 1.00 in steps of 0.05. A revelation probability of 50% is costless; any other revelation probabilities are costly. The cost function is symmetric around 50% and convex. The full cost function can be found in the instructions in the appendix. Participants next do or do not learn the color of their neighbor, with the respective revelation probability. They do not observe the revelation probability set by their neighbor. Participants then report how likable and pro-social they find their neighbor. The pro-sociality question refers to altruism in the treatments Altruism and Random, and trustworthiness in the treatment Trustworthiness. 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, regarding all parts, at the beginning of the experiment. To be precise, participants learn that they will interact with their neighbor and that the experiment has three parts. They also learn that in the last part, they determine the probability with which their neighbor might receive information on their relative pro-social behavior in the second part. The first part of the instructions is identical in Altruism and Random. Participants receive more detailed instructions directly before each part. Participants thus learn the precise signaling structure and costs of manipulating the revelation probabilities only after their pro-social decisions. Note that participants can hide their color with certainty at relatively low costs, which further alleviates ethical concerns.

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

3 Theoretical predictions

This section develops a simple theory to clarify thoughts. 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 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.

Participants might want to affect the beliefs of their neighbors by revealing their assigned color $c \in \{g, y, r\}$. Setting revelation probability p might generate costs; the cost function f is strictly convex and symmetric around 0.5 with $f(0.5)$ normalized to zero. Participants only pay to set a particular revelation probability if they believe that revealing a particular color changes the pro-social impression they make. Let $\nu(c)$ be the second-order belief of participants revealing their 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.

Consider a participant with color $c \in \{g, y, r\}$ who sets revelation probability p. Define his expected utility 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, the first-order conditions

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

defines optimal revelation probabilities $p^*(c)$ for each 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.

Note that the second-order belief $\nu(n)$ in the model—how participants in the experiment believe their neighbors to interpret not seeing any color—is a priori unclear. Not seeing any color need not be construed neutrally. For example, if people 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 hidden color must be red. Then not revealing any color is an unfavorable signal. Without further assumptions, the theory thus makes no predictions concerning the absolute probabilities with which participants might want to disclose specific colors. Moreover, not revealing any color could rationally be interpreted as an unfavorable signal, even in the treatment Random in which colors have no informational content. The theory is, therefore, consistent with participants in Random manipulating their revelation probability.

However, the crucial design component of the experiment is that colors in Altruism and Trustworthiness have a specific meaning by the rules of the experiment. In

the above simple theory, $\nu(g)$ in Altruism and Trustworthiness 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 if the prior μ is strictly between zero and one. Then $\nu(g) > \nu(y) > \nu(r)$ holds and implies $p^*(g) > p^*(y) > p^*(r)$. In Random, colors have no informational content so that $\nu(g) = \nu(y) = \nu(r) = \mu$ and $p^*(g) = p^*(y) = p^*(r)$. These predictions are independent of $\nu(n)$ as long as participants believe that how not revealing their color is interpreted by their neighbor does not depend on the hidden color. This condition seems innocuous, given that hidden colors remain private information.

Summarizing, if participants care about their social image, then they should condition their revelation probabilities on colors in Altruism and Trustworthiness. 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.

Finally, the experiment only estimates a lower bound for the importance of social image concerns for the following reasons. First, participants do not directly observe the revelation probability set by their neighbor. However, seeing or not seeing the neighbor's color does reveal some information on the revelation probability. If participants do not want to show that they care about their social image, they might want to leave the revelation probability at the cost-minimizing level. Second, participants might form strong beliefs concerning their neighbors during the face-to-face discussion. Participants who believe their social image not to be malleable should not engage in social signaling, even if they care strongly for their social reputation. Finally, participants know that information on their pro-social behavior might become observable later. If people, for example, want to avoid that others consider them to be selfish, they could behave very pro-socially. After being assigned a favorable color, there is then no need to manipulate the revelation probability.

4 Experimental results

4.1 Interpretation and informational content of colors

Table 1 shows how participants assess the pro-sociality of their neighbor, conditional on the revealed color of their neighbor. In Altruism and Trustworthiness, those who reveal more favorable colors are evaluated more favorably. Assessment and colors are uncorrelated in Random. Regression analysis confirms these impressions (p values smaller than 0.001 in Altruism and Trustworthiness, and weakly larger than 0.27 in Random, with standard errors clustered on pairs).

Non-parametric analysis shows that colors also have the intended informational content. The following bootstrapping procedure addresses the problem that observations within pairs are not independent. Within every pair, one randomly selected participant enters the statistical test, and this procedure is repeated 1000 times. Reported are the average p value and the rate at which the tests reject the null

Table 1 Assessment of pro-sociality

	Red	Yellow	Green
Altruism	2.81	3.43	4.16
Trustworthiness	2.86	4.04	4.70
Random	3.53	3.73	3.56

The table shows the average assessment of pro-sociality of the neighbor conditional on the revealed color of the neighbor. Colors reveal information on relative past donations in Altruism and past trustworthiness in Trustworthiness. Colors in Random are randomly assigned and thus have no informational content

hypothesis at the 10% significance level. Bootstrapped non-parametric tests are said to reject the null if the average p value is less than 10%. Bootstrapped Spearman rank correlation tests show that colors correlate with pro-social behavior in Altruism and Trustworthiness (average p values smaller than 0.001 with rejection rates of 1.00). Colors and pro-social behavior do not correlate in Random (average p values of 0.58 with a rejection rate of 0.03).

The signaling story requires that participants form the right beliefs about how neighbors interpret colors. The experimental design determines beliefs, and control questions ensure a common understanding of the rules. It is highly likely that the interpretation of colors is common knowledge.[2]

4.2 Revelation probabilities in Altruism and Trustworthiness

Participants who want to manipulate the pro-social impression they make on others should set a higher revelation probability if assigned a color with more favorable informational content. Figure 1 shows for each treatment the average revelation probabilities conditional on color. 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 Trustworthiness.

Table 2 shows how often revelation probabilities are below, equal, or higher than 50%, all conditional on assigned color, and for all treatments. Most participants do not pay anything to manipulate their revelation probability: 78% in Altruism and 66% in Trustworthiness leave their revelation probability at exactly 50%. However, some do pay to manipulate their revelation probability, and they behave according

[2] Although not important for identifying social image concerns, it is interesting to see how participants interpret seeing no color. The average pro-sociality assessment if no color is revealed is 3.10 in Altruism, 3.97 in Trustworthiness, and 3.34 in Random. Regression analysis confirms that in Altruism, participants consider those who reveal no color as more pro-social than those who reveal red (p value of 0.07) but as less pro-social than those who reveal yellow or green (p values of 0.02 and less than 0.001). In Trustworthiness, participants consider those who reveal no color as more pro-social than those who reveal red (p value less than 0.001), about as pro-social as those who reveal yellow (p value of 0.64) and as less pro-social than those who reveal green (p value of less than 0.001). In Random, participants consider those who reveal no color as about as pro-social as those who reveal red or green (p values of 0.27 and 0.16), and as less pro-social than those who reveal yellow (p value of 0.004).

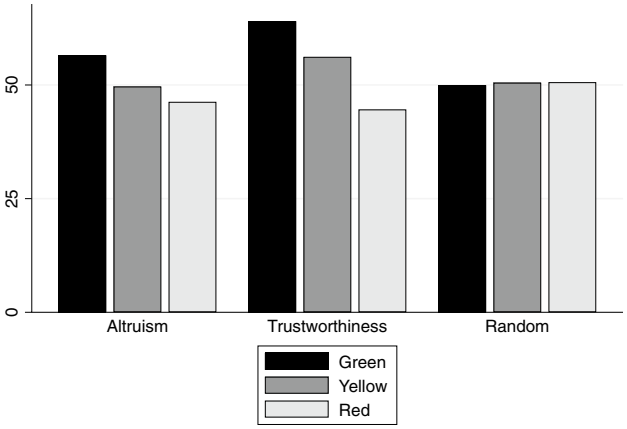

Fig. 1 Average revelation probabilities. Note: the figure shows average revelation probabilities conditional on the assigned color, and for all treatments. Colors reveal information on relative past donations in Altruism and past trustworthiness in Trustworthiness. Colors in Random are randomly assigned and thus have no informational content

Table 2 Extensive margin revelation probabilities

	Red	Yellow	Green
Altruism			
Entire sample	7–58–4	9–59–8	2–54–19
Low donations	6–55–3	5–34–4	0–10–1
High donations	1–3–1	4–25–4	2–44–18
Trustworthiness			
Entire sample	13–23–5	1–60–18	2–20–14
Low trustworthiness	12–22–3	1–30–7	0–3–1
High trustworthiness	1–1–2	0–30–11	2–17–13
Random			
Entire sample	8–60–4	6–55–7	6–61–7
Low donations	4–29–2	2–31–0	4–28–2
High donations	4–31–2	4–24–7	2–33–5

The entries in each cell of the table show how often revelation probabilities are strictly below, exactly equal, or strictly higher than 50%, conditional on the assigned color, and for all treatments. The first line in each panel takes all observations in the respective treatment. The second and third line distinguish whether the pro-social decisions made before the color assignment are above or below the median, where the categorization for Altruism and Random is based on the pooled donation data. Colors reveal information on relative past donations in Altruism and past trustworthiness in Trustworthiness. Colors in Random are randomly assigned and thus have no informational content

to theory. Table 2 also hints at the more delicate patterns of social image concerns. In Altruism, participants seemingly want to reveal having been assigned the color green because 19 participants set a high, and only 2 participants set a low revelation

probability. There is no such clear revelation pattern when assigned yellow or red—similar numbers set high and low revelation probabilities. In Trustworthiness, participants seemingly want to reveal green or yellow, and like to hide red. These interpretations are consistent with the averages reported in Fig. 1.

Table 2 also reports revelation behavior conditional on pro-social behavior. Social signaling seems to be driven by those who donate more than the median in Altruism: 18 versus 2 participants with high donations and the color green set a high rather than low revelation probability. In Trustworthiness, all participants seem to care about their social image. For example, among participants with low trustworthiness and being assigned red, 12 set a low, and 3 set a high revelation probability. Summarizing, these findings suggest that either all or mostly the more pro-social types care about their social image in the present experiment. These results contrast Friedrichsen and Engelmann (2018), who find that less pro-social types are more interested in their social image. However, participants in the current experiment know that some information about their pro-social image might become known to their neighbor. Selfish participants who care about their social image might thus behave very pro-socially. The above results should be interpreted cautiously.

The statistical analysis generates the same conclusions. Figure 2 reports the results from simple linear regressions. Baseline regresses revelation probabilities on dummies for being assigned green or red, with no further control variables. The estimated coefficients of the baseline regressions are plotted as black dots, with 95% confidence intervals clustered on pairs. They show that participants in Altruism choose on average a higher 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 smaller than 0.001). Participants in Trustworthiness choose a higher revelation probability if assigned green rather than yellow, but this effect is only weakly significant (p value of 0.08). Participants choose a lower revelation probability if assigned red rather than yellow (p value smaller than 0.001). A Wald test confirms that participants set a higher revelation probability if assigned green rather than red (p value smaller than 0.001). Conclusions are the same from ordered probit regressions in which the dependent variable indicates whether participants set a revelation probability smaller than, exactly equal to, or larger than 50%; see the numbers in Table 2. Bootstrapped Kruskal–Wallis tests come to the same conclusion (average p values of 0.04 and 0.03 with rejection rates of 0.89 and 0.92).

Social signaling might be more important in the trust game than in charitable giving. The reason is that trustworthiness measures pro-sociality towards other participants, with whom the receivers of the signal might strongly identify. Directly comparing behavior in Altruism and Trustworthiness is difficult, because participants engage in different strategic situations, which also complicates any statistical comparisons. For example, there is no reason why participants should interpret seeing no color similarly in Altruism and Trustworthiness. However, eyeballing the data suggests that social signaling is more critical for trustworthiness than donations because the estimated coefficients on red and green are further apart.

Looking at participants' self-assessment, one might expect in particular participants who are negatively surprised by the ranking feedback to set a low revelation

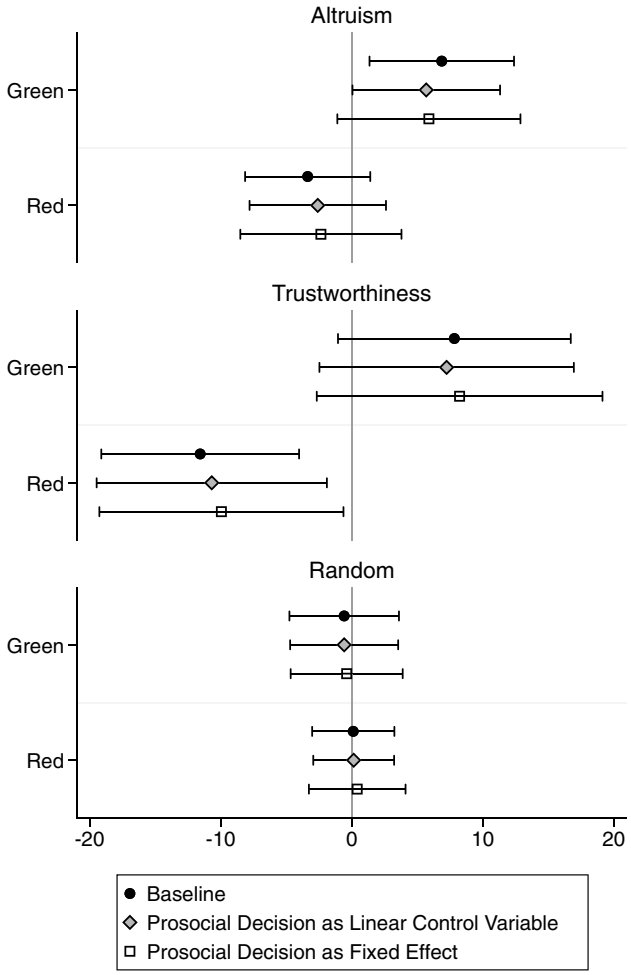

Fig. 2 Regression results. Note: the figure plots the estimated coefficients of linear regressions with revelation probability as dependent variable. Green and Red refer to dummy variables indicating the assigned color, the omitted reference category is the color yellow. The three panels refer to the treatments Altruism, Trustworthiness 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 Trustworthiness. 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 Trustworthiness, and 214 in Random

probability because the information they would reveal does not coincide with their self-image. Negatively surprised participants set low revelation probabilities in Altruism. But otherwise, neither negative nor positive surprises are correlated with revelation probabilities (p values of 0.08 and weakly larger than 0.18 in regression analyses with standard errors clustered on pairs).

Summarizing, the data suggest that at least some participants consciously manipulate their social image. The overall effects are rather weak, maybe due to the reasons discussed at the end of the theory section. Nevertheless, some participants want to manipulate the social impression they make on others, and they behave just as suggested by the theory. Capturing only a lower bound, the experiment documents the importance of social image concerns.

4.3 Endogeneity

As already discussed in the introduction, the correlation between colors and revelation probabilities might be spurious if unobservable characteristics drive both pro-social behavior and the choice of revelation probabilities. One possibility to address this concern is comparing behavior in Altruism and Random. These treatments are identical, with the only difference that colors convey no information in Random.[3] If unobserved characteristics drive donations and revelation probabilities, these variables should correlate in Altruism and Random. Bootstrapped Spearman rank correlation tests yield statistically significant correlation in Altruism (average p value of 0.09 with a rejection rate of 0.74) but not in Random (average p value of 0.44 with a rejection rate of 0.11).

Furthermore, participants who care about their social image should not condition their revelation probabilities on colors in Random, where colors have no informational content.[4] 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 Fig. 2 finds that participants in Random do not condition their revelation probability on colors (p values of at least 0.74). A bootstrapped Kruskal–Wallis test confirms this result (average p value of 0.55 with a rejection a rate of 0.05).[5]

Regression analysis is an alternative approach to address the potential endogeneity problem. The regression analysis uses the random composition of the assigned reference groups for identification. Given any pro-social behavior, the randomly assigned reference group determines the relative ranking of participants. The section next explores whether participants condition their revelation probabilities on colors once controlling for pro-social behavior.

[3] More economic students participated, and donations were lower, in Altruism rather than in Random, due to some outlier sessions. As discussed in the appendix, conclusions remain the same when controlling for the outlier sessions.

[4] Bootstrapped Spearman rank correlation tests confirm that colors in Random do not correlate with pro-social behavior (average p value of 0.58 with a rejection rate of 0.03). Regression analysis shows that colors do not affect how participants assess the pro-sociality of their neighbors (p values weakly larger than 0.27).

[5] 82% of participants in Random leave their revelation probability at the cost-minimizing 50%, almost the same as the 78% of participants in Altruism. However, spending resources on manipulating the revelation probability in Random is consistent with social image concerns; see the theory section.

Figure 2 reports the results from regressions that include pro-social decisions as a linear control variable. The estimated coefficients are plotted as grey diamonds. Participants in Altruism set a higher revelation probability if assigned green rather than yellow (p value of 0.05). A Wald test reveals that participants set a higher revelation probability if assigned green rather than red (p value of 0.01). Participants in Trustworthiness set a lower revelation probability if assigned red rather than yellow (p value of 0.02). A Wald test shows that participants set a lower revelation probability if 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 plotted as hollow squares. Controlling for the initial pro-social behavior does not affect the regression analysis very much, except for slightly inflating standard errors. Closer inspection 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 Trustworthiness. 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.

5 Conclusion

The present experiment shows that people care what you think; people strategically and thus consciously manage their social image. The findings lend further credibility to theoretical social signaling models, which typically assume that people correctly anticipate how their behavior affects their social reputation; see, for example, Austin-Smith and Fryer (2005), Bénabou and Tirole (2006) and Ellingsen and Johannesson (2008).

Social image concerns can strongly affect the efficiency of organizations; see, for example, the use of symbolic awards in Kosfeld and 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. More precise knowledge on what exactly people would like to signal might have critical managerial implications.

Acknowledgements Open Access funding provided by Projekt DEAL.

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, R. G, Jr. (2005). An economic analysis of "acting white". *Quarterly Journal of Economics*, *120*(2), 551–583.
Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*(5), 1652–1678.
Bursztyn, L., Ferman, B., Fiorin, S., Kanz, M., & Rao, G. (2018). Status goods: Experimental evidence from platinum credit cards. *Quarterly Journal of Economics*, *133*(3), 1561–1595.
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.
DellaVigna, S., List, J. A., Malmendier, U., & Rao, G. (2017). Voting to tell others. *Review of Economic Studies*, *84*, 143–181.
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.
Eriksson, T., Mao, L., & Villeval, M.-C. (2017). Saving face and group identity. *Experimental Economics*, *20*, 622–647.
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.
Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.
Friedrichsen, J., & Engelmann, D. (2018). Who cares about social image? *European Economic Review*, *110*, 61–77.
Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125.
Holm, H. J., & Samahita, M. (2018). Curating social image: Experimental evidence on the value of actions and selfies. *Journal of Economic Behavior and Organization*, *148*, 83–104.
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.
Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644.
Soetevent, A. R. (2005). Anonymity in giving in a natural context—A field experiment in 30 churches. *Journal of Public Economics*, *89*, 2301–2323.