

# Facing Yourself – A Note on Self-image

Armin Falk\*

December 3, 2018

## Abstract

The concern for a positive self-image is a central assumption in a large class of signaling models. In this paper, we exogenously vary self-image concerns by manipulating self-directed attention and study the impact on moral behavior. The choice context in the experiment is whether subjects inflict a painful electric shock on another subject to receive a monetary reward. In the main treatment, subjects see their own face on the decision screen in a real-time video feed. In three control conditions, subjects see either no video at all or a neutral video, or they see themselves in a mirror. We find that the exogenous increase in self-image concerns significantly reduces the fraction of subjects inflicting pain. The finding emphasizes the importance of self-awareness for moral decision making with implications for theory as well as practical applications to promote socially desirable outcomes.

*Keywords:* Self-image, moral behavior

*JEL Codes:* D64, C91.

---

The European Research Council (European Community's Seventh Framework Programme Grant Agreement no. 340950) provided research support. I would like to thank Jean Tirole for fruitful discussions and helpful comments, as well as Thomas Graeber for excellent research assistance.

\*Institute on Behavior and Inequality (briq) and Department of Economics, University of Bonn.

# 1 Introduction

A concern for a positive self- or social image is the central assumption of a large class of signaling models. The latter explain a variety of phenomena and behaviors such as prosociality, crowding out of motivation, will-power, norm-based behavior, taboos or notions of identity and the dual self (e.g., Bénabou and Tirole, 2006, 2011, 2013; Bodner and Prelec, 2003; Seabright, 2009; Bénabou, Falk and Tirole, 2018).<sup>1</sup> It is assumed that individuals either like to think positively of themselves or have a preference for being liked and well regarded by others. Several experimental studies have demonstrated the behavioral importance of social image,<sup>2</sup> but little is known about whether an exogenous variation in self-image affects behavior. In this paper we provide causal evidence for the behavioral importance of self-image concerns.

Essential to the concept of self-image is the human capacity of “reflexive thinking”, i.e., taking oneself as the object of attention (Leary and Tangney, 2012). In contrast to social image, self-image concerns are self-directed and refer to the awareness of congruency between individual standards and the self. To examine the effect of an exogenous variation in self-awareness, we ran an experiment with four between-subjects conditions. In our main condition, subjects see a real-time webcam video showing their face, i.e., they are confronted with their “self-image”. We compare behavior in the main condition with outcomes in three control conditions, one where subjects see no video at all, one where they see a neutral video of an unrelated other, and one with using a mirror instead of the webcam video. The choice context is moral decision-making: in the experiment, subjects face the binary decision between receiving money for inflicting a painful electric shock on another subject versus not inflicting pain and foregoing the money. This paradigm captures a widespread conception of morality according to which harming others in an intentional and unjustified way is considered immoral (Gert and Gert, 2016). Our hypothesis is derived from the signaling model of Bénabou, Falk and Tirole (2018), which assumes a lower likelihood of shocking in

---

<sup>1</sup>Relatedly, Akerlof and Kranton (2000, 2005) account for identity by positing that utility depends on the degree to which actions accord with one’s own identity. Other related signaling models that assume image concerns include Besley et al. (2015), Prendergast and Stole (2001), and Pesendorfer (1995).

<sup>2</sup>An example is Ariely et al. (2009). In their “Click for Charity” experiment, subjects donated to a charitable organization by repeatedly clicking two keys on a computer keyboard. They find that participants exert significantly greater effort in the presence of an audience than in private. Likewise, Ewers and Zimmermann (2015) show that when exposed to an audience, subjects state significantly higher quiz outcomes relative to a treatment without audience.

the main video condition relative to both control video conditions. Our data confirm this prediction, whereby an increased salience of self-image significantly reduces the likelihood of inflicting pain, i.e., the incidence of immoral behavior. This finding sheds light on self-image concerns in moral decision-making, although we suggest that it holds relevance in support of self-signaling models more generally.

A potential concern about the treatment manipulation using the webcam video is that people might mistakenly get the impression that they are also watched by someone else. To address this issue we design a control treatment in which a mirror is attached to the computer monitor, ruling out the potentially confounding influence from simultaneous social image concerns induced by the webcam. The results from the webcam manipulation replicate in the mirror condition, suggesting that the treatment effect is indeed driven by differential self-awareness.

The remainder of the paper is organized as follows. The next section discusses related work. Section 3 derives our predictions and describes the design of the experiment. Section 4 presents our results and section 5 concludes.

## 2 Related Literature

Our paper relates to empirical work in both economics and psychology, most of which invokes self-image concerns after alternative explanations fail. In this sense, several experimental investigations indirectly infer the relevance of self-image in the context of dishonesty or giving. Tonin and Vlassopoulos (2013) suggest self-image concerns as a plausible (ex-post) explanation of their finding that a fraction of subjects opt out of their preceding allocation decisions in a dictator game. Johansson-Stenman and Svedsäter (2012) document a wedge between hypothetical and real-money willingness-to-pay for moral goods as opposed to non-moral goods and interpret this as indicative of a concern for a positive moral identity. Another vibrant line of research examines the phenomenon of lying, based on a paradigm introduced by Fischbacher and Föllmi-Heusi (2013). Even though the maintenance of a positive self-concept or self-reputation is frequently entertained as a plausible explanation, most lying studies are not in and of themselves designed to pin down self-image concerns as a general mechanism underlying honest behavior. For example, Ploner and Regner (2013) explore whether behaving morally or randomly experiencing a more moral outcome in a preceding experiment subsequently affect cheating in an independent lying task. The

first-stage manipulations in Ploner and Regner (2013) are not specific to self-image concerns, however, but rather inform the more general debate about moral behavior in dynamic setups.

Other studies test behavioral implications of self-image concerns from the perspective of specific theoretical models. Mazar et al. (2008), for example, formalize self-concept maintenance in a way that people strike a balance between the preservation of a self-image of honesty and higher profits from dishonest behavior. In a series of experiments, Mazar et al. (2008) provide indirect evidence for self-concept maintenance in tasks involving honesty, by priming subjects of their moral standards – using, e.g., religious reminders – or by increasing ambiguity on how to interpret one’s actions. Grossmann (2015) tests whether giving behavior accords to the comparative statics of Bayesian signaling equilibria implied by self-signaling and social-signaling concerns. His results lend little support to an impact of self-signaling according to Bayesian equilibrium. A follow-up paper, Grossmann and van der Weele (2017), explores the specific theoretical predictions of self-image concerns in Bayesian equilibrium for settings that allow for willful ignorance. Their results conform with five general behavioral patterns implied by ignorance equilibria, which are in line with underlying self-image concerns but do not serve as a sharp experimental test thereof. Importantly, none of the above papers directly manipulates the strength of self-image concerns. A seeming exception is Cueva and Dessi (2012), who report an experiment that creates exogenous variation in the salience of self-image concerns. Before making a donation decision, the control group received an anonymous message on their screen about the preceding donation decision of another subject (the first-mover), while in the treatment group this first-mover stood up and presented a card indicating the donated amount. By altering the mode of announcing a previous subject’s decision, this design inevitably changes the salience of *social* comparison as well, through the provision of a personally visible instead of an anonymous referent. Previous work thus differs from the present experiment, which exogenously shifts self-image in terms of self-directed attention, eliminating potential confounding factors due to the presence of other individuals.

Our findings also complement recent work in economics on the general – intrinsic and extrinsic – sources of prosocial motivation (Cappelen, Halvorsen, Sørensen and Tungodden, 2016) as well as research in psychology that has stressed the importance of self-awareness (Duval and Wicklund, 1972). Related evidence shows, e.g., that self-awareness fosters fairness and honesty if moral standards are salient (Batson, Thompson, Seufferling, Whitney, and Strongman, 1999), reduces aggressive behavior (Froming et al.,

1982) and inhibits cheating in a performance test (Diener and Wallbom, 1976; Vallacher and Solodky, 1979).

### 3 Hypotheses and Design

**Theoretical background.** Models incorporating image concerns typically include a weighting parameter, reflecting the weight individuals place on social or self-image. The former captures any reputational costs and benefits coming from being observed by others and can be instrumental or affective, e.g., in terms of feelings of shame or guilt.

To illustrate the role of self-image for moral behavior in a theoretical framework, we refer to Bénabou, Falk and Tirole (2018).<sup>3</sup> In their model, an individual chooses whether to engage in moral behavior ( $a = 1$ ) or not ( $a = 0$ ). A moral decision generates an expected positive externality  $e$ , and yields a self-image benefit. The individual has deep value  $v$  (moral type) or 0 (immoral type), with probabilities  $\rho$  and  $1 - \rho$ .  $\bar{v} = \rho v$  denotes the expected value. Thus, the high type has an intrinsic motivation for the moral action equal to  $ve$ .  $c$  denotes the private cost of engaging in moral behavior and  $\beta < 1$  is a hyperbolicity parameter, measuring an individual’s potential lack of self-control. The perceived cost,  $\frac{c}{\beta}$ , is sufficiently large that the immoral type does not behave prosocially, i.e., he always chooses  $a = 0$ . The key assumption to be studied here is  $\mu$ , which measures the strength of image concerns.

In equilibrium<sup>4</sup>, the moral type chooses  $a = 1$  if and only if

$$ve - \left(\frac{c}{\beta}\right) + \mu(v - \bar{v}) > 0.$$

It immediately follows that the likelihood of moral behavior increases in  $\mu$ . Conditional on choosing  $a = 1$ , the value of self-image is given by  $\mu(v - \bar{v})$ , i.e., the value of signaling one’s morality, weighted by  $\mu$ . We interpret the weighting parameter  $\mu$  as the salience of self-image concerns. Hence,  $\mu$  is an awareness parameter, indicating a person’s attentiveness about his identity. A low  $\mu$  characterizes a decision-maker who is not attentive to learning his type, i.e., who does not pay much attention to his self-image.

---

<sup>3</sup>We derive a prediction for this model because it explicitly deals with moral decision-making for binary choice tasks, exactly as in the experiment. However, other self-signaling (see above) models deliver similar intuition and predictions.

<sup>4</sup>The following formula assumes that if there are multiple equilibria, the Pareto dominating one is selected. However, the result that an increase in image concerns increases moral behavior also holds for the alternative equilibrium selection.

Note that in this interpretation,  $\mu$  is not a hard-wired preference parameter, but rather an environmental feature, akin the presence or absence of an external audience.<sup>5</sup> It is this notion of  $\mu$  that is examined in our experiment: we exogenously increase the salience of self-image and study implications for moral behavior.

**Design of Experiment.** Studying the causal impact of self-image concerns on an individual’s moral behavior requires (i) exposing subjects to a morally-relevant decision context and (ii) to randomly vary self-image concerns.

Regarding (i), note that according to a general conception of morality, harming others in an intentional and unjustified way – especially for personal gain – is considered immoral (Gert and Gert, 2016). Informed by this notion, the decision context used in the experiment is about inflicting pain on another subject. Subjects made a binary choice between two options, labeled option A and option B. Option A implied that the subject would not receive additional money and that no other person would receive a shock. Option B implied that the subject was paid 8 euros and that another participant received a painful, yet harmless, electric shock. The instructions were specific about procedural details and informed subjects that the electric shock would be delivered with two electrodes attached to the other subject’s forearm, illustrated by a picture (see Figure 2 in the Appendix). Subjects were further informed that the shocks are medically harmless, but painful.<sup>6</sup> They also knew that the other participants would take part in another experimental session and receive a fixed payment irre-

---

<sup>5</sup>An alternative interpretation of  $\mu$  is the notion of a distinct preference about the self. An individual with low  $\mu$  might then not intrinsically value knowing or learning who he is, i.e., how his type compares to some relevant reference standard. In this case, he would show little tendency to incur a cost to signal to himself that he is a moral type. Note, however, that in the absence of self-directed attention or self-awareness, any intrinsic preference for image is unlikely to be effective. In this sense, we consider self-awareness a precondition for self-image to affect behavior. In our empirical study, we suggest that *given* an underlying preference for a positive self-image, we can manipulate the strength of this concern by directing attention toward the self. See also the discussion and references in Leary and Tangney (2012).

<sup>6</sup>Note that electric shocks are commonly used in a wide range of academic fields, in particular psychology and neurosciences (Crosbie, 1998; Mechias et al., 2010). These studies – as well as ours – are run in accordance with ethical principles in academia and are authorized by the respective ethics committees. Examples of studies using electric stimuli comprise empathy (Singer et al., 2004), anxiety-related behavior (Butler et al., 2007; Kalisch et al., 2005), neural responses to aversive stimuli (Jensen et al., 2003), operant conditioning (Crosbie, 1998; Mechias et al., 2010; Phelps, et al., 2004), and anticipatory beliefs (Falk and Zimmermann, 2017).

spective of how many shocks they received. The instructions provided little room for interpretation regarding the choice situation. Subjects were told that their decision would be about whether someone was “willing to inflict pain on another participant in return for money”.

To address (ii) – the causal effect of self-awareness – we ran four between-subjects conditions. In all conditions, participants took their decision in private, i.e., in their lab cubicle with closed curtains (see Figure 3 in the Appendix). In the main condition – “Self-image” (*SI*) – self-awareness was exogenously increased by exposing participants to seeing their face. Throughout the decision process (including the presentation of instructions and decision screens), a webcam which was placed on top of the monitor recorded their face, which was displayed through a video on the computer screen in real time (see Figures 4 and 7 in the Appendix). The device was angled and subsequently fixed in such a way that a seated subject of arbitrary height always saw the visual field of their camera. The high-resolution camera generated a clear image that captures even subtle details of facial expressions. The video screen was prominently placed in the middle-upper part of the screen and was already running when subjects entered their cubicle. At the very beginning of the session, the instructions explicitly informed subjects that the video would not be recorded or stored and that no other person aside from the subject him-/herself could view the video. To give meaning to the setup with a camera and obfuscate the experimental objective, subjects were also told that at the end of the experiment, they would be asked to answer a few short questions on the camera technology and the camera settings. None of the subjects raised questions or concerns related to the camera.

We compare decisions in *SI* to outcomes in two control treatments. In treatment “No Image” (*NoI*), everything was kept identical aside from the fact that no video was shown. The top center space used for the webcam stream in *SI* was left empty, while keeping the structure and formatting of information presented on the screens exactly identical (see Figure 6). Hence, this condition is akin to a typical lab experiment. Of course, self-image concerns may also play a role in this treatment. However, in comparison to *SI*, subjects’ attention is not explicitly drawn to themselves, arguably yielding a lower salience of self-image.

The comparison between *SI* and *NoI* may potentially be confounded for three reasons. First, subjects may simply feel distracted when seeing a video. If this distraction absorbs cognitive resources, subjects may not be able to exert the self-control necessary to inhibit selfish impulses (Gino et al., 2011;

Achtziger et al., 2015).<sup>7</sup> In this sense, distraction itself could potentially reduce the propensity to act morally. Second, seeing yourself inevitably implies seeing a human being. Previous evidence suggests that the mere fact of seeing a pair of eyes may be sufficient to induce notions of “being observed” (Burnham and Hare, 2007; Ernest-Jones et al., 2011). In other words, seeing yourself may (mistakenly) trigger “social reputational” concerns rather than enhanced self-attentiveness. To address both concerns, we ran a second control condition. In treatment “Neutral Image” (*NeuI*), subjects saw the video of a news presenter presenting news reports on German national television. In this condition, subjects may be equally distracted and see a pair of eyes, but not their own image. We chose a well-known news presenter (rather than some unknown unrelated person) to ensure that subjects immediately understood that the person in the video could not see them. Furthermore, the person shown is a non-controversial public person working for a mainstream public-service television broadcaster. As such, he does not evoke tendentious associations or is indicative of the experiment’s objectives. The video was mute and occupied the exact same place on the screen as the video in *SI* (see Figure 5). A third potential confound in the manipulation of self-awareness using a webcam is that subjects might have thought that someone else watches the video or that the video is stored, even though they were explicitly informed that this was not the case. To rule out any notion of being observed by anybody else, a third treatment condition, “Mirror Image” (*MI*) used a simple mirror, rather than a webcam. The mirror was attached to the monitor such that it occupied the same space as the video in the webcam condition (see Figure 8 in the Appendix). Moreover, the mirror was removed by the experimenter at the exact same time when the webcam video was shut down, i.e., after subjects had made their choice between options A and B.

After subjects had taken their decision, we elicited socio-demographic background characteristics together with personality-related items, in particular the Big-5 (NEO FFI 60-item version) as well as IQ (10 Raven matrices).<sup>8</sup> For participation, subjects received a show-up fee of 4 euros and an additional 8 euros if they delivered a shock to the other participant. A total of 396 subjects participated in the role of active decision-makers (46

---

<sup>7</sup>Note that if the fundamental impulse is to act prosocially rather than selfishly – as argued, e.g., in Rand et al. (2012) – distraction effects would actually bias the findings against our hypothesis.

<sup>8</sup>For this part of the experiment, the webcam in *SI* as well as the video in *NeuI* were switched off, and the mirror in *MI* was removed. Hence, in all four treatment conditions subjects were in an identical situation when answering the surveys.

percent male). Subjects were students from the University of Bonn, studying in various fields. We used z-tree as experimental software (Fischbacher, 2007). The “other participants” received a show-up fee of 20 euros for participation, irrespective of the number of shocks that they received. They were not informed about why they received the shocks. The electric shock paradigm was approved by the Ethics Committee of the University of Bonn (reference no. 156/13).<sup>9</sup>

## 4 Results

As discussed above, the model assumes that the likelihood of moral behavior increases with image concern ( $\mu$ ). Since  $\mu$  is exogenously increased in *SI* and *MI* in comparison to both *NoI* and *NeuI*, we hypothesized that the likelihood of shocking should be higher in the latter two conditions relative to the former. Indeed, this is what we find. The fraction of subjects willing to inflict pain is 0.54 (n=95) in *SI*, 0.57 (n=109) in *MI*, 0.72 (n=94) in *NoI* and 0.68 (n=98) in *NeuI*, respectively (see Figure 1). The observed effects of the self-image manipulation in *SI* relative to both no image and the neutral image are sizable and statistically significant (*SI* vs. *NoI*:  $\chi^2(1) = 7.05, p = 0.008$  and *SI* vs. *NeuI*:  $\chi^2(1) = 4.38, p = 0.036$ ; two-sided tests). There is no significant difference between the webcam and mirror implementation of self-image (*SI* vs. *MI*:  $\chi^2(1) = 0.21, p = 0.65$ ), suggesting that the webcam does not create the illusion of being watched by someone else. Note that although shocking rates are slightly lower in *NeuI* than in *NoI*, the difference is not statistically significant ( $\chi^2(1) = 0.36, p = 0.547$ ). Thus, simply seeing the face of an unrelated third person is insufficient to effectively lower immoral behavior relative to not seeing anyone.

Table 1 reports OLS regressions<sup>10</sup> where we regress the decision to inflict pain on the other subject on three treatment dummies (*NoI*, *NeuI* and *MI*) with *SI* as the omitted category.

Column (1) shows the raw treatment effects. Subjects are significantly less willing to inflict pain when facing themselves (in treatments *SI* or *MI*) compared to not seeing themselves or seeing a neutral video, respectively. Shocking frequencies are not significantly different between treatments *NoI* and *NeuI*, as shown in the footer of Table 1. In column (2), we include age and gender. Column (3) additionally controls for other socio-demographic (income) and personality-related (Big-5 and IQ) information and shows that

<sup>9</sup>For instructions of the experiment and further details, see the Appendix.

<sup>10</sup>Probit estimates yield the same results.

Figure 1: Fraction of subjects inflicting pain

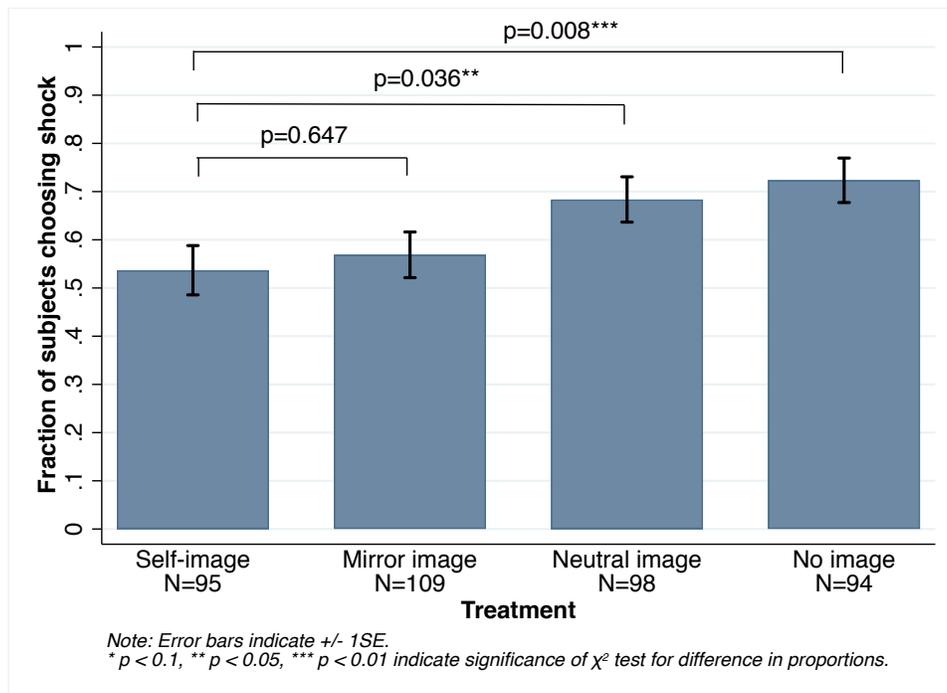


Table 1: OLS Regressions on Full Sample

	Dependent variable: Shocking choice			
	(1)	(2)	(3)	(4)
No image	0.187*** (0.0692)	0.180*** (0.0671)	0.162** (0.0668)	0.166** (0.0670)
Neutral image	0.147** (0.0698)	0.156** (0.0665)	0.162** (0.0644)	0.158** (0.0638)
Mirror image	0.0320 (0.0701)	0.0554 (0.0677)	0.0489 (0.0657)	0.0525 (0.0655)
Gender (1=male)		0.265*** (0.0463)	0.222*** (0.0510)	0.223*** (0.0507)
Age		-0.00137 (0.00563)	-0.00370 (0.00534)	-0.00323 (0.00526)
Cognitive intelligence (Raven)			-0.0351*** (0.0135)	-0.0347** (0.0135)
Available income			0.000137 (0.000107)	0.000140 (0.000108)
Big5: Neuroticism			0.0192 (0.0270)	0.0189 (0.0266)
Big5: Extraversion			0.0670** (0.0261)	0.0654** (0.0259)
Big5: Agreeableness			-0.113*** (0.0222)	-0.111*** (0.0219)
Big5: Openness			-0.0493** (0.0228)	-0.0532** (0.0229)
Big5: Conscientiousness			0.0104 (0.0262)	0.00683 (0.0259)
Estimated pain from shock				0.0384* (0.0215)
Constant	0.537*** (0.0514)	0.442*** (0.144)	0.672*** (0.163)	0.462** (0.205)
No=Neutral (F)	0.360	0.141	0.00000268	0.0168
No=Neutral (p)	0.549	0.707	0.999	0.897
F	3.421	9.036	10.50	10.30
N	396	396	396	396

Robust standard errors in parantheses.

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

our treatment effects are robust to adding these control variables. Female participants as well as those scoring high on IQ and agreeableness (one of the five facets of the Big-5 inventory) are significantly less likely to inflict pain, which is consistent with findings of related work on moral behavior (e.g., Deckers et al., 2016; Falk and Szech, 2013).

As part of the post-experimental questionnaire, we asked subjects about their belief regarding how painful it is to receive an electric shock, i.e., about the perceived externality. This belief was elicited in an incentive-compatible way. To do so, we used a sample of 24 subjects who had previously received a series of shocks in an unrelated study and who were asked to state how painful they actually had experienced receiving these shocks on a seven-point Likert scale. In the present experiment, we explained this to subjects and asked them to indicate on the same seven-point scale which number is closest to the mean rating of these other subjects. A correct answer was remunerated with 2 euros. Including these estimates does not change our main result (see column (4)). Moreover, the belief about the painfulness of the shock is not significantly correlated with the decision to inflict pain (the raw correlation across all treatments is  $\rho = 0.072$  ( $p = 0.152, n = 396$ )). One possible explanation is that despite having a material incentive to tell the truth, subjects who inflicted pain may have shifted beliefs in a self-serving manner.

## 5 Concluding remarks

We have shown that an exogenous increase in the salience of self-image reduces immoral behavior. This finding has several implications. Most importantly, it lends empirical support to the assumption of self-image concerns, the foundation of a large class of self-signaling models. Individuals not only care about social image, but also about a positive image vis-à-vis themselves. On a practical level, our findings suggest new mechanisms and instruments to promote morally desirable outcomes. Firms, organizations or tax authorities seeking to promote socially responsible behaviors may want to create environments that draw individuals' attention to themselves when taking decisions. For example, forms and contracts could be designed to include pictures, require personal signatures or ask the decision-maker to reflect on who he/she is or would like to be.

In our experiment, we have primarily been interested in moral decision-making. However, we believe that the video paradigm used to manipulate the level of self-image concerns could also prove useful in assessing the role of

self-image in other choice contexts. This essentially applies to all situations where people like to think positively about themselves. Examples are self-control problems, lying, rule-following or norm compliance (Abeler et al., forthcoming; Kimbrough and Vostroknutov, 2016). We speculate that when facing themselves, individuals will display higher levels of self-control (e.g., less myopic discounting, sticking to plans, change eating or drug habits), act more honestly, and are more likely to obey rules. This suggests further important practical applications to reduce the incidence of unwanted behaviors.

A further extension that can easily be implemented using the video paradigm involves endogenizing self-awareness, i.e., subjects could be allowed to choose whether or not they want to be exposed to seeing themselves. We would expect that the likelihood of actively avoiding self-awareness is higher if decisions are image-relevant and costly (like costly moral decision-making, as in our study) in comparison to decision contexts that are not costly or relevant for self-image. If the salience of self-image can in fact be managed, this would suggest further behavioral implications such as the active avoidance of contexts and situations that remind individuals of themselves, or that are associated with personal experiences or memories.

## References

- Abeler, J., D. Nosenzo and C. Raymond (forthcoming) "Preferences for Truth-Telling," *Econometrica*.
- Achtziger, A., Alós-Ferrer, C., and A. K. Wagner (2015) "Money, Depletion, and Prosociality in the Dictator Game," *Journal of Neuroscience, Psychology, and Economics*, 8(1): 1–14.
- Akerlof, George A., and Rachel E. Kranton (2000) "Economics and Identity," *Quarterly Journal of Economics*, 115(3): 715-753.
- Akerlof, George A., and Rachel E. Kranton (2005) "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, 19(1): 9-32.
- Ariely, D., Bracha, A. and S. Meier (2009) "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *American Economic Review*, 99(1): 544-555.
- Batson, C. D., Thompson, E. R., Seufferling, G., Whitney, H. and J.A. Strongman (1999) "Moral Hypocrisy: Appearing Moral to Oneself Without Being So," *Journal of Personality and Social Psychology*, 77(3): 525-537.
- Baumeister, R. F. (1998) "The Self", In Gilbert, D., Fiske, S. T., and G. Lindzey (Eds.), *The handbook of social psychology* (pp. 680-740). New York: Oxford University Press.
- Bénabou, R., A. Falk and J. Tirole (2018) "Narratives, Imperatives and Moral Reasoning," *NBER Working Paper No. 24798*.
- Bénabou, R. and J. Tirole (2006) "Incentives and Prosocial Behavior," *American Economic Review*, 96(5): 1652-1678.
- (2011) "Identity, Morals and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, 126(2): 805-855.
- (2013) "Laws and Norms," *Mimeo*, Toulouse School of Economics.
- Besley, T., Jensen, A., and T. Persson (2015) "Norms, Enforcement, and Tax Evasion," *CEPR Discussion Papers*, No. 10372.

- Bodner, R., and D. Prelec (2003) "Self-signaling and diagnostic utility in everyday decision making," *The Psychology of Economic Decisions*, 1: 105-26.
- Burnham, T. C. and B. Hare (2007) "Engineering Human Cooperation: Does Involuntary Neural Activation Increase Public Goods Contributions?," *Human Nature*, 18: 88-108.
- Butler, T., Pan, H., Tuescher, O., Engelién, A., Goldstein, M., Epstein, J., ... and L. Chang (2007) "Human Fear-Related Motor Neurocircuitry," *Neuroscience*, 150(1): 1-7.
- Cappelen, A., T. Halvorsen, E. Sørensen and B. Tungodden (2016) "Face-Saving or Fair-Minded: What Motivates Moral Behavior?," *Journal of the European Economic Association* (in press).
- Crosbie, J. (1998) "Negative Reinforcement and Punishment," *Handbook of research methods in human operant behavior*, Springer US, 1998. 163-189.
- Cueva, C., and R. Dessi (2012) "Charitable giving, self-image and personality" *Institut d'Economie Industrielle (IDEI), Toulouse*, No. 748'.
- Deckers, T., Armin, F., Kosse, F., and N. Szech (2016) "Homo moralis: Personal characteristics, institutions, and moral decision-making," *IZA Discussion Paper*, No. 9768.
- Diener, E. and M. Wallbom (1976) "Effects of Self-Awareness on Antinormative Behavior," *Journal of Research in Personality*, 10(1): 107-111.
- Duval, S. and R. A. Wicklund (1972) *A Theory of Objective Self Awareness*. Oxford: Academic Press.
- Ernest-Jones, M., Nettle, D., and M. Bateson (2011) "Effects of Eye Images on Everyday Cooperative Behavior: A Field Experiment," *Evolution and Human Behavior*, 32: 172-178.
- Ewers, M. and F. Zimmermann (2015) "Image and Misreporting," *Journal of the European Economic Association*, 12(2): 363-380.
- Falk, A., and N. Szech (2013) "Morals and Markets," *Science*, 340(6133): 707-711.

- Falk, A. and F. Zimmermann (2017) “Beliefs and Utility: Experimental Evidence on Preferences for Information,” *Discussion paper, University of Bonn*.
- Fischbacher, U. (2007) “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments,” *Experimental economics*, 10(2): 171-178.
- Fischbacher, U., and F. Föllmi-Heusi (2013) “Lies in Disguise – An Experimental Study on Cheating,” *Journal of the European Economic Association*, 11: 525-547.
- Froming, W. J., G. R. Walker, and K. J. Lopyan (1982) “Public and Private Self-Awareness: When Personal Attitudes Conflict with Societal Expectations,” *Journal of Experimental Social Psychology*, 18(5): 476-487.
- Gert, B. and J. Gert (2016) “The Definition of Morality,” *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/spr2016/entries/morality-definition/>
- Gino, F., Schweitzer, M. E., Mead, N. L. and D. Ariely (2011) “Unable to Resist Temptation: How Self-control Depletion Promotes Unethical Behavior,” *Organizational Behavior and Human Decision Processes*, 115(2): 191-203.
- Grossman, Z. (2015) “Self-signaling and social-signaling in giving,” *Journal of Economic Behavior and Organization*, 117: 26-39.
- Grossman, Z., and J. van der Weele (2017) “Self-image and strategic ignorance in moral dilemmas,” *Journal of the European Economic Association*, 15(1): 173-217.
- James, W. (1890) *The Principles of Psychology*. New York: Holt.
- Jensen, J., McIntosh, A. R., Crawley, A. P., Mikulis, D. J., Remington, G., and S. Kapur (2003) “Direct Activation of the Ventral Striatum in Anticipation of Aversive Stimuli,” *Neuron*, 40(6): 1251-1257.
- Johansson-Stenman, O., and H. Svedsäter (2012) “Self-image and valuation of moral goods: Stated versus actual willingness to pay,” *Journal of Economic Behavior and Organization*, 84(3): 879-891.

- Kalisch, R., Wiech, K., Critchley, H. D., Seymour, B., O’doherly, J. P., Oakley, D. A., ... and R. J. Dolann (2005) “Anxiety Reduction Through Detachment: Subjective, Physiological, and Neural Effects,” *Journal of cognitive neuroscience*, 17(6): 874-883.
- Kimbrough, E. O., and A. Vostroknutov (2016) “Norms make preferences social,” *Journal of the European Economic Association*, 14(3): 608-638.
- Leary, M. R. and J. P. Tangney (2012) *Handbook of Self and Identity, Chapter 1: The Self as an Organizing Construct*, (2nd ed.). New York: The Guilford Press.
- Mazar, N., Amir, O., and D. Ariely (2008) “The dishonesty of honest people: A theory of self-concept maintenance,” *Journal of Marketing Research*, 45(6): 633-644.
- Mechias, M.-L., Etkin, A., and R. Kalisch (2010) “A Meta-Analysis of Instructed Fear Studies: Implications for Conscious Appraisal of Threat,” *Neuroimage*, 49(2): 1760-1768.
- Pesendorfer, W. (1995) “Design Innovation and Fashion Cycles,” *American Economic Review*, 85(4): 771-792.
- Phelps, E. A., Delgado, M. R., Nearing, K. I., and J. E. LeDoux (2004) “Extinction Learning in Humans: Role of the Amygdala and vmPFC,” *Neuron*, 43(6): 897-905.
- Ploner, M., and T. Regner (2013) “Self-image and moral balancing: An experimental analysis,” *Journal of Economic Behavior and Organization*, 93: 374-383.
- Prendergast, C., and L. Stole (2001) “The Non-Monetary Nature of Gifts,” *European Economic Review*, 45: 1793-1811.
- Rand, D. G., Greene, J. D. and M. A. Nowak (2012) “Spontaneous Giving and Calculated Greed,” *Nature*, 489(7416): 427-430.
- Seabright, P. B. (2009) “Continuous Preferences and Discontinuous Choices: How Altruists Respond to Incentives,” *The BE Journal of Theoretical Economics*, 9(1): 1935-1704.
- Singer, T., Seymour, B., O’doherly, J., Kaube, H., Dolan, R. J., and C. D. Frith (2004) “Empathy for Pain Involves the affective but Not Sensory Components of Pain,” *Science*, 303(5661): 1157-1162.

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

Vallacher, R. R. and M. Solodky (1979) “Objective Self-Awareness, Standards of Evaluation, and Moral Behavior,” *Journal of Experimental Social Psychology*, 15(3): 254-262.

## A Instructions (Computer-Based)

*Instructions are translated from German into English.*

**Welcome screen.** Instructions: Thank you for your participation. For your participation, you will receive 4 euros. The money will be paid to you in cash at the end. Please note: During the entire experiment, communication between participants is forbidden. Please only use the designated functions on your computer. If you have questions, please make a hand signal. Your question will be answered at your seat.

All statements in these instructions are true. This holds generally for all experiments at the Bonn Laboratory for Experimental Economics Research, as well as for this experiment. In particular, all consequences of actions that are described in the instructions will be executed exactly as described.

You can earn additional money depending on how you decide. In addition, your decisions have consequences for another participant in a different experiment.

**Camera instructions screen** (*treatment SI only*). Information on camera video: As you can see, a camera has been attached to the monitor.

Please note: No recordings are saved, and only you and no other person sees the camera video. At the end of the experiment, we will ask you a few short questions on the camera technology and the camera settings.

**Mood screen** On a scale from 0 to 10, how is your current mood? Please indicate your answer on the scale, where 0 means “very bad” and 10 “very good”.

**Shocking choice instructions, main screen.** Your choice: In the following, you have the choice between Option A and Option B.

If you choose Option A, you will receive no additional money.

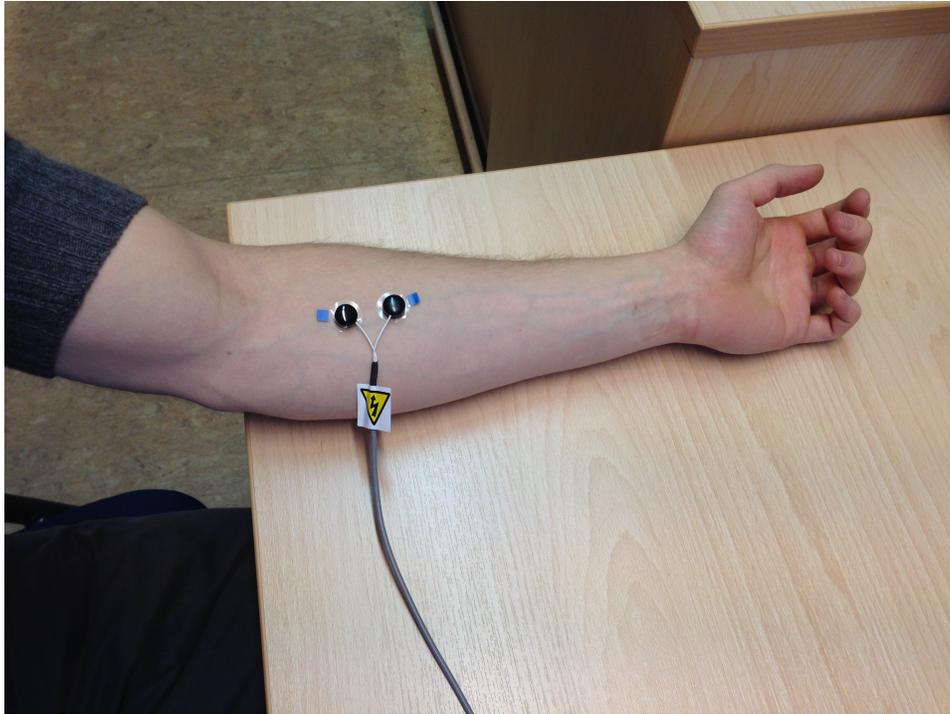
If you choose Option B, you will receive an additional amount of 8 euros.

Your decision has a further consequence. If you choose Option B, another participant in a different experiment will receive a painful electric shock. The impulse is administered using two electrodes, which are attached to the forearm of the participant (see picture to the right [see Figure 2]). The electric impulses are harmless to health, but very painful.

Your decision in this experiment is therefore whether you are willing to in-

flict pain on someone in return for money.

Figure 2: Picture of shocking device shown to participants



**Shocking choice instructions, details screen.** As already mentioned, you can choose one of two options. Once you have decided for one option, this option will be executed. If you chose Option A, you will not receive additional money and the other participant will receive no electric shock. If you chose Option B, you will receive an additional amount of 8 euros at the end of the experiment and the other participant will receive a painful electric shock.

Please note: The other participant takes part in a different experiment at a different date. The other participant will be informed that he/she will potentially receive a painful electric shock in this experiment, and he/she gives his written consent to participation in the experiment according to the guidelines of the Ethics Committee. The other participant receives money for his/her participation in the experiment, and this is independent of whether

he/she receives an electric shock or not.

Anonymity: You will not know the identity of the other participant at any point in time, and your identity is also completely anonymous.

**Shocking choice instructions, summary screen.** Summary: In Option A, you will not receive an additional payment, and the other participant will receive no painful electric shock. In Option B, you will receive an additional amount of money, and the other participant will receive a painful electric shock. The decision is yours.

You will make your choice on a decision screen, which will appear soon.

**Pre decision screen.** On the next screen, you can now choose between Options A and B.

**Decision screen.** Please decide now.

I choose

Option A

Option B

Confirm

Figure 3: Lab with closed curtains, ensuring full privacy



Figure 4: Decision screen in treatment *SI* with camera video (this figure is blurred for privacy reasons, but subjects saw their face in high resolution)

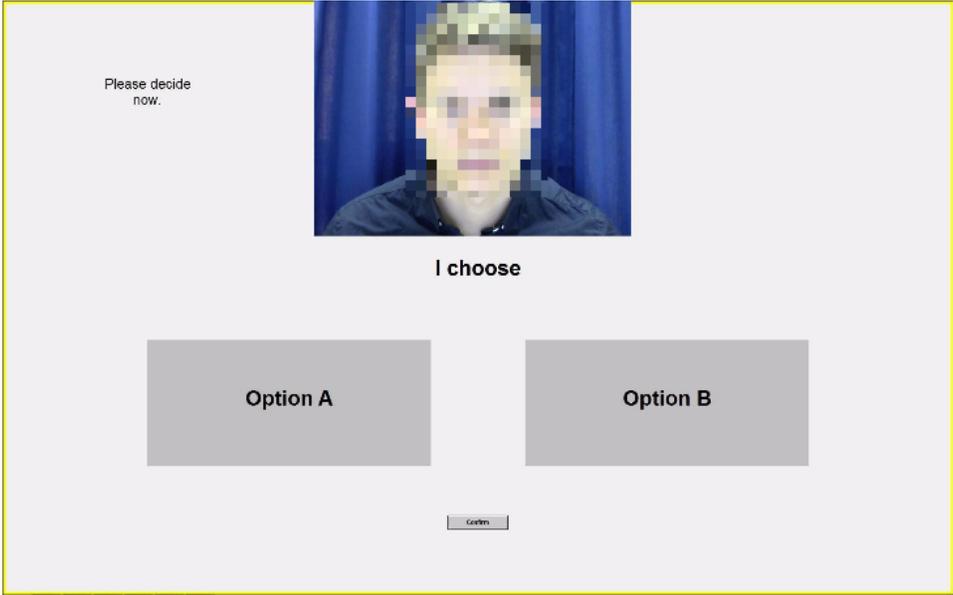


Figure 5: Decision screen in treatment *NeuI* with neutral video

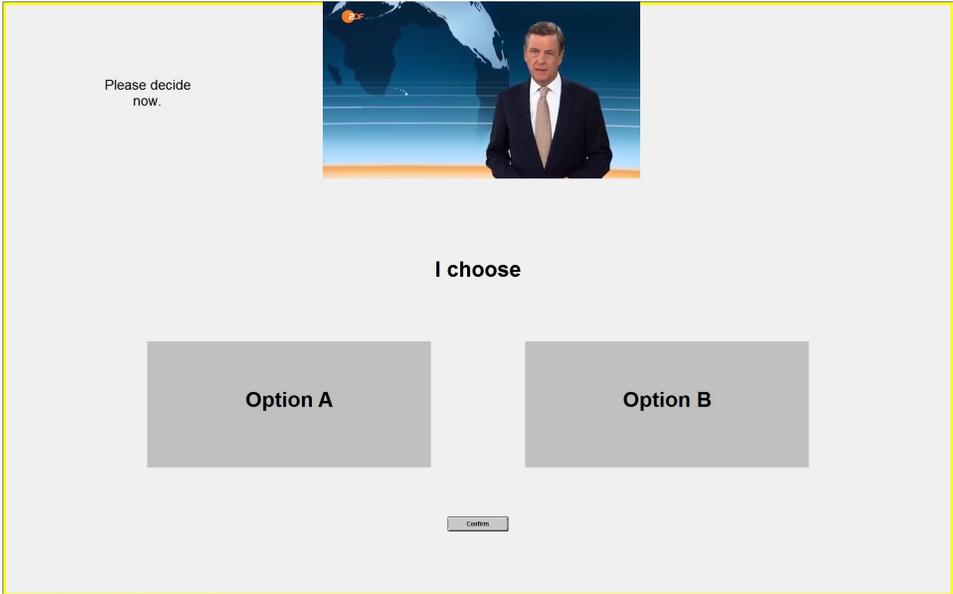


Figure 6: Decision screen in treatment *NoI* without any video

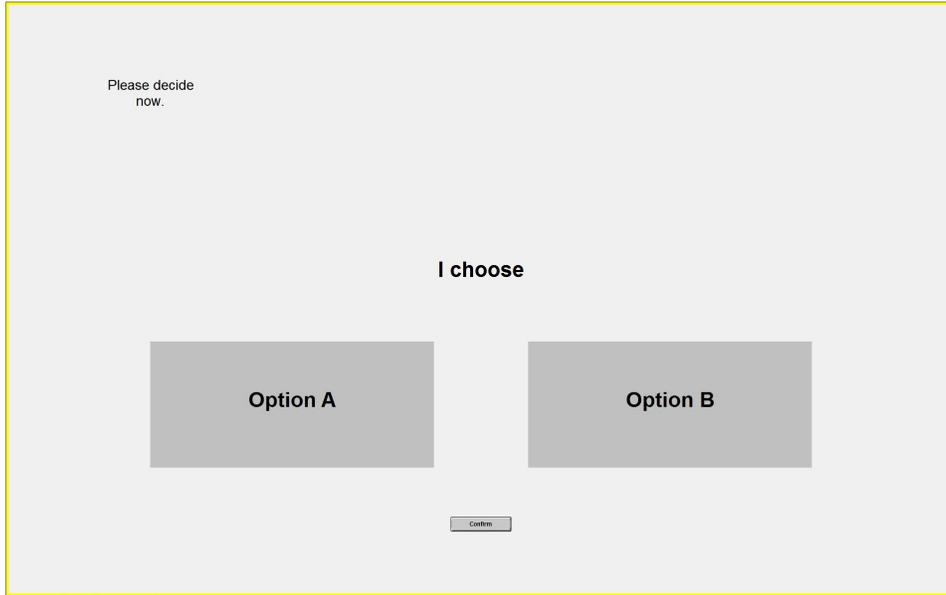


Figure 7: Laboratory cubicle with camera attached to monitor in treatment *SI*

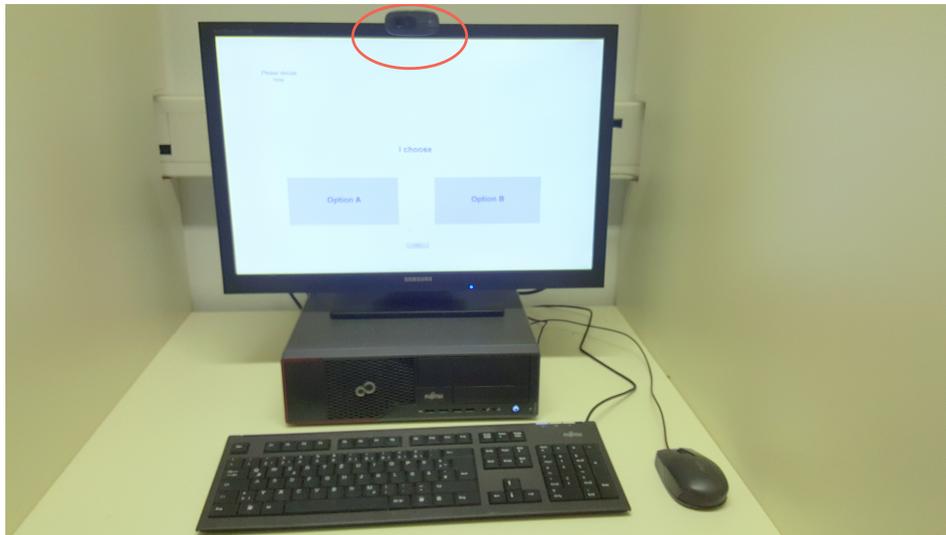


Figure 8: Laboratory cubicle with mirror attached to monitor in treatment *MI*

