Essays on Image Concerns and Norm-Enforcing Behavior

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Zvonimir Bašić

aus Našice, Kroatien

Bonn, 2019

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Armin Falk
Zweitreferent:	Prof. Dr. Sebastian Kube
Tag der mündlichen Prüfung:	13. Dezember 2018

Acknowledgments

First and foremost, I'd like to deeply thank Armin Falk for all of his support and all the interesting and intellectually stimulating discussions that we had during my doctoral journey. His rigorous approach and out-of-the-box thinking has strongly molded the way I approach science, and will surely resonate through my work in the future. Moreover, I would like to thank Sebastian Kube for his feedbacks, discussions and support in multiple aspects over the years, especially for his support when I was deciding where to pursue my doctoral studies. Finally, I would thank Pia Pinger for her valuable and honest feedbacks and her participation as the third committee member on my thesis defense.

Next, I would like to thank all the colleagues and friends providing the intellectually stimulating environment that shaped my thoughts over the last 5 years. In particular, Fabian Kosse for teaching me all the relevant basics (and much more) at the beginning of my doctoral studies; Simone Quercia for hundreds of discussions and the valuable knowledge; Jonas Radbruch, whose excess of knowledge fit perfectly with my excess of motivation, resulting in long study sessions during my master studies which were crucial for my transition towards doctoral studies; Dmitry Kuvshinov for the support, the discussions, and making sure to always stay positive and have a laugh; Felix Albrecht, Peter Andre, Markus Antony, Holger Gerhardt, Roberto González, Adrian Hillenbrand, Lukas Kießling, Andrea Reykers, Stefanie Sauter, Sebastian Schaube, Felix Schran, Louis Strang, Matthias Wibral, Jana Willrodt, Kaspar Zimmermann, and many others.

I would also like to thank all the people at the Max Planck Institute in Bonn for an amazing last year, and making sure that my "final stretch" of the doctoral thesis goes smooth. In particular, Matthias Sutter for offering me the amazing position at the institute, and numerous friends and colleagues such as Stefania Bortolotti, Zita Coppinger, Zwetelina Iliewa, Sofia Monteiro, Matthias Praxmarer, Shambhavi Priyam, Angelo Romano, Daniel Salicath, Ali Seyhun Saral, Sebastian Schneider, Eugenio Verrina, Claudia Zoller and others.

Finally, I would like to thank my parents Tomislav and Ksenija and my sister Ivana who were always an emotional support and a strong motive to push further in my studies, and who made sure that, despite not living in ideal times, I had an amazing early life environment. Also, I would like to thank my close friends from Croatia and from my master studies for their support over the last years, and my further family from Slavonija, most of all, grandma Rozika and aunt Truda, and my late grandparents Fina, Joso and Pavo, who determined a great deal of who I am today. Finally, and most importantly, I would like to give my biggest gratitude to my girlfriend Alina. She sacrificed way too much for my

thesis, and was always there when things got tough. I literally cannot imagine how any of this would have been possible without her endless support and love.

Contents

List of Figures

List of Tables

1

Introduction

Economic decisions reveal information about the person making them. In plenty of situations, they are informative of some important and valued traits, for instance, of one's prosociality or morality. This gives rise to a strong behavioral motive, as people care about how the society perceives them, but also, they care about how they perceive themselves. These two motives, commonly known as *social and self-image concerns* (see Bénabou and Tirole, 2006), are frequently present in our day to day life, and thus have a propensity to shape a variety of our decisions. Moreover, people do not only care about what others might *think* of them due to their behavior, but they also are concerned about what others might *do* if the behavior is met with disapproval. In particular, violation of important social norms is often followed by *norm-enforcing behavior*, as others punish those that violate them (see Fehr and Fischbacher, 2004). Similarly to image concerns, the threat of punishment is a common motive, as the society is intertwined with relevant social norms which people are willing to enforce.

This thesis consists of four essays which contribute to a better understanding of *the behavioral impact of image concerns* as well as *norm-enforcing behavior*. All of them employ experiments, either in a lab setting or a lab-in-the-field setting, and together focus on the behavior of adults as well as the behavior of children and adolescents.

In Chapter 2 (joint work with Armin Falk and Simone Quercia), I study how self and social image concerns influence prosocial behavior. A stream of evidence from economics and psychology indicates the importance of the two notions of image concerns as drivers of prosocial behavior; however, no study compares them. We, therefore, develop a symmetric design that allows us to contrast the influence of the two notions. In a large scale dictator game, we exogenously increase self-awareness and observability in order to direct subjects' focus on their private and public self, respectively. We show that both self and social image concerns are drivers of prosocial behavior. We observe, however, that manipulating observability causes a stronger increase in prosocial behavior compared to

self-awareness. We also document a marked gender difference in self-image concerns, shedding light on the different effect sizes across our treatments. While both genders react similarly to the observability manipulation, only men react to the self-awareness manipulation, closing the initial gender gap in prosocial behavior. We report evidence indicating that men tend to be generally less self-aware than women, i.e., they are less focused on their prosocial standards, but can be "nudged" towards them.

In Chapter 3 (joint work with Armin Falk and Simone Quercia), I study how self and social image concerns influence lying behavior. Image concerns are often hypothesized as an important component of lying costs. In a standard die-rolling paradigm, we expose subjects to the same manipulations as in Chapter 2, and find that an increase in self-awareness has no effect on their reports. In contrast, we show that an increase in subjects' observability, while still maintaining their private information, significantly decreases their reports (and with it, their profits). We finally show in a survey experiment that respondents believe that the likelihood of a lie increases with the reported outcome, and attribute negative traits to people who make high reports. This further supports reputation concerns as the explanation behind the results of our social image treatment.

In Chapter 4 (joint work with Armin Falk and Simone Quercia), I study how self and social image concerns influence prosocial behavior in childhood and adolescence. While lots of studies focus on the effect of image concerns in adults, very little is known about their roots in young age. We show that in a dictator game with 7-14 year-old subjects, both self and social image concerns matter, however, only for boys: i) boys give more when social image concerns are increased, significantly more so than girls, who do not react, ii) boys give more when self-image concerns are increased, while girls do not, and iii) supporting evidence indicates that the observed gender asymmetry is not due to differences in understanding of normative behavior, suggesting an actual difference in concerns for appearing prosocial. Our results support evolutionary theories suggesting that men should be more inclined to signal prosociality as well as recent predictions about the age when reputation should start playing a role.

In Chapter 5 (joint work with Armin Falk and Fabian Kosse), I study the development of egalitarian norm enforcement in childhood and adolescence. Egalitarian norm is a long-existing organizing principle, commonly enforced by adults across societies. We investigate the development of this enforcing-behavior with 9-18 year-old subjects, by taking the most commonly-used third-party punishment game where a third party is added to a dictator game, and adapting it for children. We show that children start enforcing the egalitarian norm at the age of 11-12. In addition, we show that: i) as the egalitarian norm enforcement emerges, a non-negligible proportion of punishers also disapprove of overly-generous transfers that exceed the norm, ii) the punishers' behavior only changes until 13-14 years of age, indicating that egalitarian norm enforce-

ment is mainly developed by that period, and iii) the dictators increase their transfer in the direction of the egalitarian norm primarily in the period when norm enforcement develops.

Taken together, this thesis focuses on the behavioral impact of self and social image concerns and norm-enforcing behavior. It explores the two by employing several experimental paradigms and eliciting behavior across different age cohorts, thus providing a multitude of insights and contributions to the respective literatures.

References

Bénabou, Roland and Jean Tirole (2006): "Incentives and prosocial behavior." *The American Economic Review*, 96 (5), 1652–1678. [1]

Fehr, Ernst and Urs Fischbacher (2004): "Third-party punishment and social norms." *Evolution and Human Behavior*, 25 (2), 63–87. [1]

2

Self-image, Social Image and Prosocial Behavior *

Joint with Armin Falk and Simone Quercia

2.1 Introduction

Prosocial behavior is widespread in humans, and it influences a wide range of societal functions and outcomes, such as the provision of public goods, redistribution, charitable giving, volunteering, contract enforcement, trade, and growth (Fehr and Gächter, 2002; Knack and Keefer, 1997; DellaVigna et al., 2012; Carpenter and Myers, 2010; Fehr et al., 1997; Guiso et al., 2009; Zak and Knack, 2001). Given its relevance, a large literature is dedicated to understanding its underlying motives. In recent years, increasing attention has been given to *self-image* and *social image concerns* (Bénabou and Tirole, 2006; Dana et al., 2007; Ariely et al., 2009; DellaVigna et al., 2012; A. Gneezy et al., 2012; Grossman, 2014; Falk, 2017; Grossman and van der Weele, 2017; Bénabou et al., 2018). Self-image concerns refer to an individual's desire to view him- or herself as prosocial. Social image concerns, on the other hand, refer to an individual's desire to be viewed as prosocial by others. Both motives are considered as crucial determinants of prosociality, and are especially interesting given their potential as basis for cost-effective policies (Andreoni and Petrie, 2004; Alpizar et al., 2008; Powell et al., 2012; Shu et al., 2012). For example, putting emphasis on one's social image was shown to induce a 25% to 48% increase in donations (Alpizar et al., 2008; Powell et al., 2012). Given that in 2016 in the US alone

* We thank Peter Andre, Dorothea Kübler, Johanna Mollerstrom, Angelo Romano, Peter Schwardmann, Egon Tripodi, Roberto Weber, Joël van der Weele, and various seminar audiences for helpful comments. We gratefully acknowledge financial support by Leibniz Programme of the German Science Foundation (DFG) and Institute on Behavior & Inequality (briq).

donations exceeded 390 billion dollars, such policies can have a substantial economic impact.

While there exists evidence on the importance of both self-image (Dana et al., 2007; Grossman, 2014; Grossman and van der Weele, 2017; Falk, 2017) and social image concerns (Andreoni and Bernheim, 2009; Ariely et al., 2009; DellaVigna et al., 2012), no study simultaneously investigates and compares the two domains. Therefore, in this paper, we design symmetric manipulations which directly and independently test the influence of self-image and social image concerns on prosocial behavior, allowing us to directly compare the two image domains. In addition, we investigate the effect of gender within our paradigm, as evolutionary theories conjecture that gender had a strong mediating role in the evolution of costly prosocial signaling (Amotz Zahavi and Avishag Zahavi, 1999; E. A. Smith and Bird, 2000; Barclay, 2010; Böhm and Regner, 2013; Van Vugt and Iredale, 2013; Raihani and S. Smith, 2015).

To design our manipulations, we build on recent theoretical work of Bénabou et al. (2018), which implies that self-image concerns arise from the awareness of discrepancy between internal standards or values and the self (e.g., in light of current behavior). Following this framework, we propose that an increase in one's *self-awareness* will make self-image concerns more salient and cause a larger personal cost when deviating from internal standards. This reasoning dates back to the theory of objective self-awareness (Duval and Wicklund, 1972), which is supported by a stream of psychological studies (Duval and Wicklund, 1973; Carver, 1975; Diener and Wallbom, 1976; Duval et al., 1979; Beaman et al., 1979). Regarding social image concerns, Bénabou et al. (2018) conceptualize them as reputational concerns vis-à-vis others, similarly to other studies (Bénabou and Tirole, 2006, 2011). Reputation has attracted the interest of several disciplines, since it represents a key evolutionary pillar of cooperation and prosocial behavior in humans (Roberts, 1998; Nowak and Sigmund, 1998; Panchanathan and Boyd, 2004). In light of this, to make social image concerns more salient we put participants' reputation at stake by increasing the *observability* of their actions.

We test the influence of the two image domains on generosity, a fundamental facet of human prosociality. In all of our treatments, participants ($n = 531$) play a dictator game where they have to decide how much of their 10 euro endowment they would like to give to a recipient who participates in a different experimental session. To manipulate self-awareness we expose participants to a real-time video of their own face on the computer screen, which is transmitted from a video camera installed on top of the computer screen as in Falk (2017). To manipulate observability, we expose participants to a real-time video of another participant's face, in an analogous way as in the self-image manipulation. This other participant is also sitting in a private cubicle in the lab, and is viewing the decision maker's face and his or her computer screen in real time. It is gener-

ally difficult to compare the relative strength of self and social image concerns, and previous ways to manipulate the two differ typically in many dimensions (Alpizar et al., 2008; Andreoni and Bernheim, 2009; Ariely et al., 2009; DellaVigna et al., 2012; Grossman, 2014; Falk, 2017; Grossman and van der Weele, 2017). We render such a comparison plausible by using the same type of manipulation, a video, varying only whether participants see themselves or are being seen by someone else. We compare these two treatments to a *Control* treatment where participants see a neutral prerecorded video of another person. The prerecorded video ensures that the effects of our image treatments are not due to cognitive depletion (Rand et al., 2012; Schulz et al., 2014; Achtziger et al., 2015; Rand et al., 2016) or social cues (Haley and Fessler, 2005; Bateson et al., 2006; Rigdon et al., 2009, see Materials and methods section (2.4) for detailed descriptions of treatments and procedures).

2.2 Results

Fig. 2.1 shows the average euro amount given by dictators across our treatments. We first compare the amount given by dictators in our *Control* treatment ($n = 188$) with the one in the *Self-image* treatment ($n = 189$). Dictators in the *Control* treatment gave on average 1.54 euros, while they gave 1.93 euros in the *Self-image* treatment. This difference is significant at the 10% level (two-sided t-test, $p = 0.065$, Cohen's $d = 0.191$). Moving to the *Social image* treatment ($n = 154$), we observe a much stronger impact of the manipulation on prosocial behavior, as participants on average gave 2.49 euros, in contrast to 1.54 euros in the *Control* treatment (two-sided t-test, $p < 0.001$, Cohen's $d = 0.475$). The average amount given in the *Social image* treatment was also significantly larger than the one in the *Self-image* treatment (two-sided t-test, $p = 0.013$, Cohen's $d = 0.270$). These results are robust to controls for socio-demographics and Big 5 personality traits using regression analyses (see Table 2.1 in Appendix).

Next, we investigate treatment differences across genders. Gender is of particular interest in our setting as it has been proposed that evolution favored asymmetric concerns for social image and costly signaling across genders. Specifically, men might benefit more from signaling generous traits in public as a successful mating strategy compared to women (Amotz Zahavi and Avishag Zahavi, 1999; E. A. Smith and Bird, 2000; Barclay, 2010; Böhm and Regner, 2013; Van Vugt and Iredale, 2013; Raihani and S. Smith, 2015). Figure 2.2 depicts our treatment effects separately for men (left panel) and women (right panel). In line with the stylized facts from previous literature (Engel, 2011), we observe that in the *Control* treatment women give on average significantly more than men (1.95 vs. 0.94 euros; two-sided t-test, $p < 0.001$, Cohen's $d = 0.539$). However, when comparing the amount given by men and women in the *Self-image* treat-

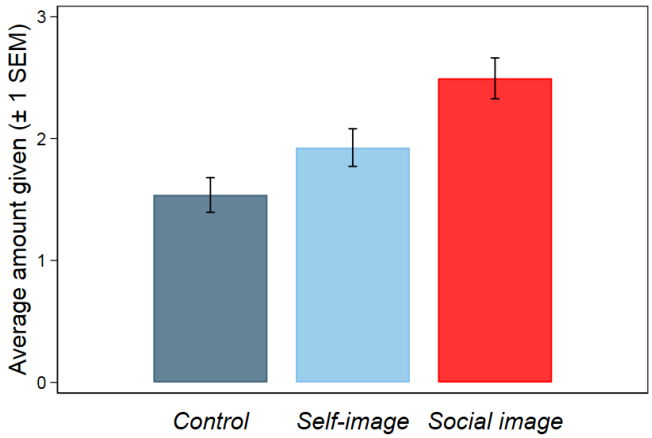

Figure 2.1. Average euro amount given by dictators across treatments

Notes: Error bars indicate standard error of the means.

ment, the gender gap disappears. In particular, women give on average 2.04, while men give 1.76 (two-sided t-test, $p = 0.388$, Cohen's $d = 0.128$). This difference is primarily due to a change in men's behavior. Men's giving increases significantly by 0.82 euros, while women's giving only increases by 0.09 euros, respectively (two-sided t-test, $p = 0.016$ for men, $p = 0.758$ for women, Cohen's $d = 0.395$ for men, Cohen's $d = 0.041$ for women). Regression analysis confirms that men react significantly stronger to the *Self-image* treatment compared to women (see Table 2.2 in Appendix). Next, we look at the effect of our social image manipulation across genders. Similar to the *Control* treatment, there exists a gender gap in giving, albeit smaller, in the *Social image* treatment. Men give 2.10 euros on average, while women give 2.75 (two-sided t-test, $p = 0.056$, Cohen's $d = 0.317$). The persistence of this gap is largely driven by the fact that both men and women increase the amount given when exposed to higher observability. Men increase their giving by 1.16 euros, while women increase their giving by 0.80 euros (two-sided t-test, $p = 0.001$ for men, $p = 0.004$ for women, Cohen's $d = 0.577$ for men, Cohen's $d = 0.411$ for women). Regression analysis confirms that there is no significant difference across genders in the reaction to increased observability (see Table 2.2 in Appendix). Comparing the two image treatments, we observe no difference for men, but a significant difference for women (two-sided t-test, $p = 0.390$ for men, $p = 0.010$ for women, Cohen's $d = 0.148$ for men, Cohen's $d = 0.362$ for women). The interaction between gender and the two image treatments is, however, not significant (see Table 2.2 in Appendix). Finally, we find no effect of the observer's gender on the average amount given by men or women (see Table 2.3 in Appendix).

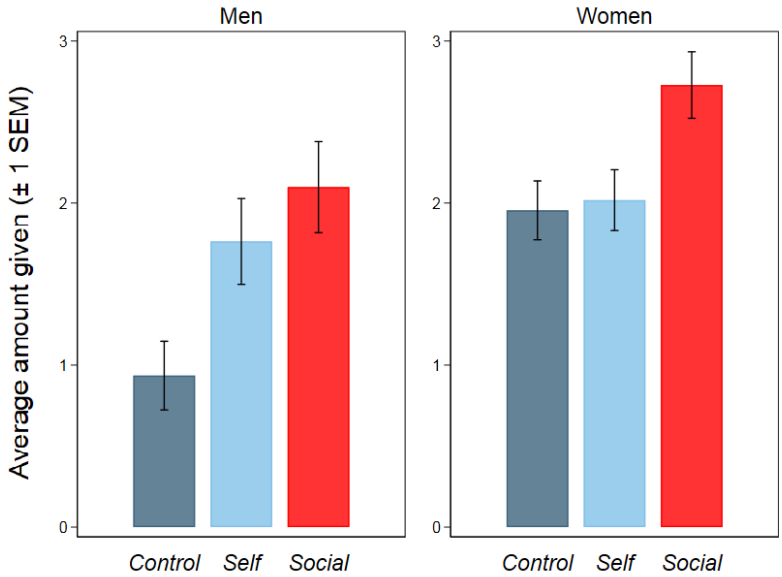

Figure 2.2. Average euro amount given by male dictators (left panel) and female dictators (right panel) across treatments

Notes: Error bars indicate standard error of the means.

The gender results shed light on the weak overall effect of our *Self-image* treatment, and raise an important question on the potential underlying mechanisms of the observed self-image gender difference. It is difficult to reconcile this finding in light of aforementioned evolutionary theories, as they primarily suggest a gender difference in the social image domain, for which we find only weak evidence. Alternatively, a straightforward hypothesis might be that women are naturally more self-aware than men, i.e., they are generally more aware of their inner standards of prosociality, while men need to be "reminded" of them. We report three measures supporting this hypothesis. First, we investigate the level of situational self-awareness across genders in the *Self-image* and *Control* treatment. For this purpose, we elicited self-focus using a sentence completion task, as self-focus is indicative of the current level of self-awareness (Exner Jr., 1973; Carver and Scheier, 1978, see Subsection 2.4.5 for a detailed description). The elicitation occurred after the decisions were made but before the cameras were turned off. We test whether women are more self-focused in the *Control* treatment but not in the *Self-image* treatment. As expected, we find that women exhibit significantly higher self-focus compared to men only in the *Control* treatment (two-sided t-test, $p = 0.039$ for *Control* treatment, $p = 0.334$ for *Self-image* treatment). Next, we check if there is a dispositional tendency for women to be more self-aware. We invited a subsample of our main sample

one week after the main experiment ($n = 207$, see Subsection 2.4.6 for details), in order to collect the measures of interest when our camera manipulation had arguably no effect anymore. We first test whether women are more privately self-conscious than men. Private self-consciousness measures a consistent tendency to focus one's attention towards oneself, i.e., it indicates the level of "chronic self-awareness" (Fenigstein et al., 1975; Carver and Scheier, 1978, see Subsection 2.4.6 for a detailed description). We find evidence that, indeed, women are more privately self-conscious than men (two-sided t-test, $p = 0.057$). Furthermore, if women have a tendency of being more aware of their inner standards, one might expect them to perceive those standards as more relevant for themselves compared to men. We analyze the relevance of prosocial values as a part of one's self-image by eliciting the importance of generosity for self-image (Grossman and van der Weele, 2017, see Subsection 2.4.6 for a detailed description). As expected, we find that women score significantly higher than men (two-sided t-test, $p < 0.001$). Overall, this evidence suggests that women may be in general more self-aware about their inner standards; hence, self-image concerns may play an important role for them even in the absence of our treatment manipulation. It further suggests that men's inner standards are also prosocial, but men need to be "nudged" towards them.

2.3 Discussion and conclusion

We have shown that both self and social image concerns are important drivers of prosocial behavior. Building on a recent theoretical literature, we have designed exogenous manipulations increasing self-awareness and observability to study the influence of self and social image concerns, respectively. One big advantage of our design is that it allows drawing parallels between the two image domains and opens the discussion about the relative importance of self and social image concerns. Across our comparable manipulations, the one affecting social image induces a stronger increase in prosociality, suggesting that being concerned about one's own social reputation may on average be a more powerful motive than being concerned about one's own opinion of oneself. Our findings offer relevant implications in light of potential policies, as both self and social image concerns can be cost-effective ways of inducing a person to act more in line with normative standards (Andreoni and Petrie, 2004; Alpizar et al., 2008; Powell et al., 2012; Shu et al., 2012). For instance, to increase the amount of charitable donations, one could ask specific personal questions to raise self-awareness (Silvia and Eichstaedt, 2004), or make donations public to raise observability. Our results indicate that both dimensions of image-based policies can yield a desirable increase in donations, but they favor an increase in observability as the more effective approach.

Interestingly, the weaker effect of the *Self-image* treatment can be explained by the fact that only men react to the increase of self-awareness. This finding contributes to the literature on gender and prosociality, where women are generally found to be more prosocial (Andreoni and Vesterlund, 2001; Croson and U. Gneezy, 2009; Engel, 2011; Falk et al., forthcoming). Recent studies propose that this difference may be due to a distinction on a neurobiological level (Soutschek et al., 2017), or due to a gender specific disposition to intuitive decision-making which affects prosocial behavior (Rand et al., 2016). We offer another piece of evidence showing that men might in general be less self-aware than women. We have provided evidence along these lines using three different measures that all support the explanation that men are less focused on their inner standards. Reminding them of their "inner self", however, corrects for this difference and shows that their inner standards are also prosocial, but not as salient as for women. This opens interesting questions for future research. Our self-awareness evidence would suggest that similar gender findings should also exist in other morally relevant domains. If so, such studies could illuminate gender differences on a broader level.

2.4 Materials and methods

2.4.1 Subject pool

Participants were recruited from the BonnEconLab subject pool with hroot (Bock et al., 2014). 531 subjects participated as dictators and were matched with recipients that participated in a different experimental session. For the purposes of the *Social image* treatment, additional 154 subjects were recruited for the role of observers. Out of the 531 dictators, 317 were women (59.7%), and the mean age was 23.78 ($sd = 5.69$).

2.4.2 Experimental procedures

The experiment was conducted in the BonnEconLab at the University of Bonn. The sessions were ran in February 2016 and July 2017. Participants were seated in cubicles and were given several minutes to read the dictator game instructions alone. After that, the experimenter read the instructions aloud. Participants were informed that their respective recipients did not participate in the current session, but would participate in the study at a different point in time. They were also informed that the recipients will be fully informed about the game structure and their decision in the game (for experimental instructions see Appendix 2.F). Before taking their decisions, participants had to answer several control questions (see Appendix 2.B). After having taken their decisions, participants were exposed to additional tasks and a questionnaire. Subjects did not

know any details of the subsequent tasks while they were taking decisions in the dictator game. The experiment was conducted using z-tree software (Fischbacher, 2007).

2.4.3 Treatments

The experiment consisted of 3 treatments: *Self-image*, *Social image* and *Control*. In all treatments dictators were facing identical procedures, except for the content of the video on their decision screen and information explaining the video. In the *Self-image* treatment, they saw a real-time video of their face transmitted from a camera installed on top of their computer screen (see Figure 2.3 and Figure 2.4 in Appendix). In the *Social image* treatment, they saw a real-time video of their paired observer's face streamed from a camera installed on the observer's computer screen. Likewise, observers saw a real-time video of the dictator's face streamed from a camera installed on the dictator's computer screen. Moreover, the observers also saw a real-time image of the computer screen of the decision maker. In the *Control* treatment, participants saw a prerecorded video of a famous German news presenter during a news segment (see Figure 2.5 in Appendix). The videos were already on at the moment participants entered the cubicles, and were turned off before the questionnaire began. In the *Self-image* treatment, participants were informed that only they could see the video, and that the video was not being recorded. In the *Social image* treatment, the roles of dictators and observers were randomly assigned. Observers had a special set of paper instructions laid in front of their computer, which were not read aloud. These instructions explained that their only task was to observe their paired dictator and that they would earn 8 euros for participation. Dictators were aware about the observer's task. Furthermore, they knew that only their paired observer could see their real-time video, that the video was not being recorded, and that the observers would leave the lab before them. They were not informed of the observer's payoff to avoid any type of social comparison, e.g., inequity aversion (Fehr and Schmidt, 1999). In both the *Self* and *Social image* treatment, in order to give some meaning to the camera, subjects were also informed that at the end of the experiment, they would answer a few short questions on the camera technology and settings. In the *Control* treatment, we used the video of the news presenter rather than a condition with no video for two main reasons: first, it allows us to control for potential cognitive depletion (Rand et al., 2012; Schulz et al., 2014; Achtziger et al., 2015; Rand et al., 2016) due to the mere presence of a video, and second, it controls for potential social cues effects that could arise due to the presence of a "pair of eyes" (Haley and Fessler, 2005; Bateson et al., 2006; Rigdon et al., 2009). Given the choice of our *Control* treatment, both the video and the "pair of eyes" are present in all three treatments, thus eliminating the potential confounds. As the content

of the video was immediately recognizable, participants were perfectly aware that the video was prerecorded. Also, the presenter's neutrality ensured that no tendentious associations were triggered.

2.4.4 Post-experimental questionnaire

The post-experimental questionnaire comprised socio-demographics questions, 60-item NEO Five Factor Inventory (Costa and McCrae, 1992) and, in the *Self* and *Social image* treatments, a few short questions about the camera settings and technology. Furthermore, in the *Social image* treatment, dictators were asked if they knew their observer before participating. Only one participant reported of knowing the paired observer. All results are unchanged if we exclude this observation.

2.4.5 Self-focus

We measured self-focus as a proxy for self-awareness. Specifically, self-focus and self-awareness are tightly connected notions, and self-focus is indicative of the current level of self-awareness (see Carver and Scheier, 1978). For measuring self-focus, we implemented the sentence completion task, which was previously used in the psychology literature on objective self-awareness (Exner Jr., 1973; Carver and Scheier, 1978). The task consists of 30 stems of sentences that need to be completed. For example, the first three sentences are: "I think...", "I was happiest when...", and "It's fun to daydream about..." (see Appendix 2.C for the list of all the stem sentences). To categorize the answers we follow Exner Jr. (1973), that is, two research assistants unaware of our research project independently coded the answers based on the following four categories: self-focused, external world focused, ambivalent or neutral answer. The answers were assigned to one of the four categories only if both research assistants indicated the same category (73% of all answers). The self-focus variable represents the frequency of self-focused answers. The task was ran after the decisions but before the cameras were turned off, and was implemented only for the sessions conducted in July 2017 ($n = 93$ for *Control* treatment, $n = 95$ for *Self-image* treatment).

2.4.6 Private self-consciousness and the importance of generosity for self-image

To elicit private self-consciousness, we used the self-consciousness scale (Fenigstein et al., 1975; Carver and Scheier, 1978). This scale consists of 23 statements. Participants had to indicate to what extent each statement would apply to them on a five-point likert scale (see Appendix 2.D for the list of the statements). To estimate the importance of prosocial values as a part of one's self-image,

we elicited the importance of generosity for self-image (Grossman and van der Weele, 2017), which is an adaptation of the self-importance of moral identity questionnaire (Aquino and Reed II., 2002). We applied the scale specifically in the context of generosity given that participants played a dictator game. Participants were asked to imagine a person with the characteristic of being generous. Subsequently, they were prompted to indicate agreement or disagreement with six statements about the importance of this attribute to their sense of self on a six-point likert scale (see Appendix 2.E for the list of statements). The measure was generated by summing the six scores. To elicit private self-consciousness and the importance of generosity for self-image, we invited dictators participating in July 2017 sessions to participate in an additional session which took place one week after the main experiment. 80% ($n = 207$) of the participants of the first session came to the session one week after (see Appendix 2.A for robustness checks on attrition). Participants were informed of the second week session only at the end of the first week session, thus preserving full comparability between sessions ran in February 2016 and July 2017.

References

Achtziger, Anja, Carlos Alós-Ferrer, and Alexander K. Wagner (2015): "Money, depletion, and prosociality in the dictator game." *Journal of Neuroscience, Psychology, and Economics*, 8 (1), 1–14. [7, 12]

Alpizar, Francisco, Fredrik Carlsson, and Olof Johansson-Stenman (2008): "Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica." *Journal of Public Economics*, 92 (5-6), 1047–1060. [5, 7, 10]

Andreoni, James and B. Douglas Bernheim (2009): "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects." *Econometrica*, 77 (5), 1607–1636. [6, 7]

Andreoni, James and Ragan Petrie (2004): "Public goods experiments without confidentiality: A glimpse into fund-raising." *Journal of Public Economics*, 88 (7-8), 1605–1623. [5, 10]

Andreoni, James and Lise Vesterlund (2001): "Which is the fair sex? Gender differences in altruism." *The Quarterly Journal of Economics*, 116 (1), 293–312. [11]

Aquino, Karl and Americus Reed II. (2002): "The self-importance of moral identity." *Journal of Personality and Social Psychology*, 83 (6), 1423–1440. [14]

Ariely, Dan, Anat Bracha, and Stephan Meier (2009): "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *The American Economic Review*, 99 (1), 544–555. [5–7]

Barclay, Pat (2010): "Altruism as a courtship display: Some effects of third-party generosity on audience perceptions." *British Journal of Psychology*, 101 (1), 123–135. [6, 7]

Bateson, Melissa, Daniel Nettle, and Gilbert Roberts (2006): "Cues of being watched enhance cooperation in a real-world setting." *Biology Letters*, 2 (3), 412–414. [7, 12]

Beaman, Arthur L., Bonnel Klentz, Edward Diener, and Soren Svanum (1979): "Self-awareness and transgression in children: Two field studies." *Journal of Personality and Social Psychology*, 37 (10), 1835–1846. [6]

Bénabou, Roland, Armin Falk, and Jean Tirole (2018): "Narratives, imperatives and moral reasoning." *Working Paper*. [5, 6]

Bénabou, Roland and Jean Tirole (2006): "Incentives and prosocial behavior." *The American Economic Review*, 96 (5), 1652–1678. [5, 6]

Bénabou, Roland and Jean Tirole (2011): "Identity, morals, and taboos: Beliefs as assets." *The Quarterly Journal of Economics*, 126 (2), 805–855. [6]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014): "hroot: Hamburg registration and organization online tool." *European Economic Review*, 71, 117–120. [11]

Böhm, Robert and Tobias Regner (2013): "Charitable giving among females and males: An empirical test of the competitive altruism hypothesis." *Journal of Bioeconomics*, 15 (3), 251–267. [6, 7]

Carpenter, Jeffrey and Caitlin Knowles Myers (2010): "Why volunteer? Evidence on the role of altruism, image, and incentives." *Journal of Public Economics*, 94 (11-12), 911–920. [5]

Carver, Charles S. (1975): "Physical aggression as a function of objective self-awareness and attitudes toward punishment." *Journal of Experimental Social Psychology*, 11 (6), 510–519. [6]

Carver, Charles S. and Michael F. Scheier (1978): "Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence." *Journal of Personality and Social Psychology*, 36 (3), 324–332. [9, 10, 13]

Costa, Paul T. and Robert R. McCrae (1992): *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources. [13]

Croson, Rachel and Uri Gneezy (2009): "Gender differences in preferences." *Journal of Economic Literature*, 47 (2), 448–474. [11]

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang (2007): "Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness." *Economic Theory*, 33 (1), 67–80. [5, 6]

DellaVigna, Stefano, John A. List, and Ulrike Malmendier (2012): "Testing for altruism and social pressure in charitable giving." *The Quarterly Journal of Economics*, 127 (1), 1–56. [5–7]

Diener, Edward and Mark Wallbom (1976): "Effects of self-awareness on antinormative behavior." *Journal of Research in Personality*, 10 (1), 107–111. [6]

Duval, Shelley, Virginia H. Duval, and Robert Neely (1979): "Self-focus, felt responsibility, and helping behavior." *Journal of Personality and Social Psychology*, 37 (10), 1769–1779. [6]

Duval, Shelley and Robert A. Wicklund (1972): *A theory of objective self awareness*. Oxford: Academic Press. [6]

Duval, Shelley and Robert A. Wicklund (1973): "Effects of objective self-awareness on attribution of causality." *Journal of Experimental Social Psychology*, 9 (1), 17–31. [6]

Engel, Christoph (2011): "Dictator games: A meta study." *Experimental Economics*, 14 (4), 583–610. [7, 11]

Exner Jr., John E. (1973): "The self focus sentence completion: A study of egocentricity." *Journal of Personality Assessment*, 37 (5), 437–455. [9, 13]

Falk, Armin (2017): "Facing yourself – A note on self-image." *CESifo Working Paper Series* (6428). [5–7]

Falk, Armin, Anke Becker, Thomas J. Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (forthcoming): "Global evidence on economic preferences." *Quarterly Journal of Economics*. [11]

Fehr, Ernst and Simon Gächter (2002): "Altruistic punishment in humans." *Nature*, 415 (6868), 137–140. [5]

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger (1997): "Reciprocity as a contract enforcement device: Experimental evidence." *Econometrica*, 65 (4), 833–860. [5]

Fehr, Ernst and Klaus M. Schmidt (1999): "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114 (3), 817–868. [12]

Fenigstein, Allan, Michael F. Scheier, and Arnold H. Buss (1975): "Public and private self-consciousness: Assessment and theory." *Journal of Consulting and Clinical Psychology*, 43 (4), 522–527. [10, 13]

Fischbacher, Urs (2007): "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10 (2), 171–178. [12]

Gneezy, Ayelet, Uri Gneezy, Gerhard Riener, and Leif D. Nelson (2012): "Pay-what-you-want, identity, and self-signaling in markets." *Proceedings of the National Academy of Sciences*, 109 (19), 7236–7240. [5]

Grossman, Zachary (2014): "Strategic ignorance and the robustness of social preferences." *Management Science*, 60 (11), 2659–2665. [5–7]

Grossman, Zachary and Joël van der Weele (2017): "Self-image and willful ignorance in social decisions." *Journal of the European Economic Association*, 15 (1), 173–217. [5–7, 10, 14]

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2009): "Cultural biases in economic exchange?" *The Quarterly Journal of Economics*, 124 (3), 1095–1131. [5]

Haley, Kevin J. and Daniel M. T. Fessler (2005): "Nobody's watching?: Subtle cues affect generosity in an anonymous economic game." *Evolution and Human behavior*, 26 (3), 245–256. [7, 12]

Knack, Stephen and Philip Keefer (1997): "Does social capital have an economic payoff? A cross-country investigation." *The Quarterly Journal of Economics*, 112 (4), 1251–1288. [5]

Nowak, Martin A. and Karl Sigmund (1998): "Evolution of indirect reciprocity by image scoring." *Nature*, 393 (6685), 573–577. [6]

Panchanathan, Karthik and Robert Boyd (2004): "Indirect reciprocity can stabilize cooperation without the second-order free rider problem." *Nature*, 432 (7016), 499–502. [6]

Powell, Kate L., Gilbert Roberts, and Daniel Nettle (2012): "Eye images increase charitable donations: Evidence from an opportunistic field experiment in a supermarket." *Ethology*, 118 (11), 1096–1101. [5, 10]

Raihani, Nichola J. and Sarah Smith (2015): "Competitive helping in online giving." *Current Biology*, 25 (9), 1183–1186. [6, 7]

Rand, David G., Victoria L. Brescoll, Jim A. C. Everett, Valerio Capraro, and Hélène Barcelo (2016): "Social heuristics and social roles: Intuition favors altruism for women but not for men." *Journal of Experimental Psychology: General*, 145 (4), 389–396. [7, 11, 12]

Rand, David G., Joshua D. Greene, and Martin A. Nowak (2012): "Spontaneous giving and calculated greed." *Nature*, 489, 427–430. [7, 12]

Rigdon, Mary, Keiko Ishii, Motoki Watabe, and Shinobu Kitayama (2009): "Minimal social cues in the dictator game." *Journal of Economic Psychology*, 30 (3), 358–367. [7, 12]

Roberts, Gilbert (1998): "Competitive altruism: From reciprocity to the handicap principle." *Proceedings of the Royal Society of London B: Biological Sciences*, 265 (1394), 427–431. [6]

Schmitt, David P., Anu Realo, Martin Voracek, and Jüri Allik (2008): "Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures." *Journal of Personality and Social Psychology*, 94 (1), 168–182. [26]

Schulz, Jonathan F., Urs Fischbacher, Christian Thöni, and Verena Utikal (2014): "Affect and fairness: Dictator games under cognitive load." *Journal of Economic Psychology*, 41, 77–87. [7, 12]

Shu, Lisa L., Nina Mazar, Francesca Gino, Dan Ariely, and Max H. Bazerman (2012): "Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end." *Proceedings of the National Academy of Sciences*, 109 (38), 15197–15200. [5, 10]

Silvia, Paul J. and Jan Eichstaedt (2004): "A self-novelty manipulation of self-focused attention for internet and laboratory experiments." *Behavior Research Methods, Instruments, & Computers*, 36 (2), 325–330. [10]

Smith, Eric Alden and Rebecca L. Bliege Bird (2000): "Turtle hunting and tombstone opening: Public generosity as costly signaling." *Evolution and Human Behavior*, 21 (4), 245–261. [6, 7]

Soutschek, Alexander, Christopher J. Burke, Anjali Raja Beharelle, Robert Schreiber, Susanna C. Weber, Iliana I. Karipidis, Jolien ten Velden, Bernd Weber, Helene Haker, Tobias Kalenscher, and Philippe N. Tobler (2017): "The dopaminergic reward system underpins gender differences in social preferences." *Nature Human Behaviour*, 1 (11), 819–827. [11]

Van Vugt, Mark and Wendy Iredale (2013): "Men behaving nicely: Public goods as peacock tails." *British Journal of Psychology*, 104 (1), 3–13. [6, 7]

Zahavi, Amotz and Avishag Zahavi (1999): *The handicap principle: A missing piece of Darwin's puzzle*. Oxford University Press. [6, 7]

Zak, Paul J. and Stephen Knack (2001): "Trust and growth." *The Economic Journal*, 111 (470), 295–321. [5]

Appendix 2.A Second week's attendance and attrition on observables

To measure private self-consciousness and the importance of generosity for self-image, dictators from our July 2017 sessions were reinvited at the end of the session to participate one week after the main experiment in an additional session. To match observations from the first to the second week and still preserve anonymity, participants were instructed to generate a code which consisted of the first two numbers of their month of birth, the first two letters of their mother's maiden name, and the first two letters of their place of birth. 80% of the subjects who were reinvited showed up in the second week. To control for selection on observables, in Table 2.4, we estimate whether variables collected in the first week can be a predictor of attendance in the second week's session. We test the effect of the treatment assignment, dictator game behavior, socio-demographics and personality characteristics. None of the variables has a significant impact on the probability of attendance in the second week.

Appendix 2.B Control questions

Please answer the following questions:[1]

1. How many euros do you receive, and how many euros does participant B receive at the beginning of the study?
2. Can participant B influence your decision?
3. If you send 4 euros to participant B, how many euros do you keep?
4. If you send 2 euros to participant B, how many euros do you keep?

[1] The control questions were administered to dictators in all treatments.

Appendix 2.C Sentence completion task

In the following part of the experiment your task is to complete unfinished sentences. The beginning of the sentence is given, for example: "The weather...". You have to complete the sentence on the "..." position. Please write whatever comes to your mind first, without long thinking.

1. I think...
2. I was happiest when...
3. It's fun to daydream about...
4. My father...
5. If only I could...
6. It's hardest for me...
7. I wish...
8. As a child I...
9. I am...
10. I'm at my best...
11. Others...
12. When I look in the mirror...
13. If only I would...
14. At least I'm not...
15. My sex life...
16. It upsets me when...
17. The thing I like best about myself...
18. Friends...
19. I would like most to be photographed...
20. I guess I'm...
21. My mother...
22. I wonder...
23. The worst thing about me...
24. I always wanted...
25. I try hardest to please...
26. Someday I...
27. My appearance...
28. My parents...
29. If I had my way...
30. I like...

Appendix 2.D Self-consciousness scale

In the following you will see statements that might apply to you. Please answer on a scale from 1 to 5 how these statements apply to you. 1 means "it does not apply to me at all" and 5 means "it applies to me fully".

1. I'm always trying to figure myself out.[2]
2. I'm concerned about my style of doing things.
3. Generally, I'm not very aware of myself. [2]
4. It takes me time to overcome my shyness in new situations.
5. I reflect about myself a lot. [2]
6. I'm concerned about the way I present myself.
7. I'm often the subject of my own fantasies. [2]
8. I have trouble working when someone is watching me.
9. I never scrutinize myself.[2]
10. I get embarrassed very easily.
11. I'm self-conscious about the way I look.
12. I don't find it hard to talk to strangers.
13. I'm generally attentive to my inner feelings. [2]
14. I usually worry about making a good impression.
15. I'm constantly examining my motives.[2]
16. I feel anxious when I speak in front of a group.
17. One of the last things I do before I leave my house is look in the mirror.
18. I sometimes have the feeling that I'm off somewhere watching myself.[2]
19. I'm concerned about what other people think of me.
20. I'm alert to changes in my mood.[2]
21. I'm usually aware of my appearance.
22. I'm aware of the way my mind works when I work through a problem.[2]
23. Large groups make me nervous.

[2] Only the marked statements are used to estimate private self-consciousness. In order to calculate the private self-consciousness variable, the answers on statements 3 and 9 were reversed, and all the marked answers were added up.

Appendix 2.E The importance of generosity for self-image

Here is one characteristic that could describe a person: generous. The person with this characteristic could be you or it could be someone else. For a moment, visualize in your mind the kind of person who has this characteristic. Imagine how that person would think, feel, and act.

When you have a clear image of what this person would be like, please respond to the statements below by indicating how strongly you agree or disagree. For each statement, the responses are (from left to right): strongly disagree, disagree, slightly disagree, slightly agree, agree, strongly agree.

Choose the response that best expresses your feelings.

1. It would make me feel good to be a person who has this characteristic.
2. Being someone who has this characteristic is an important part of who I am.
3. A big part of my emotional well-being is tied up in having this characteristic.
4. Having this characteristic is an important part of my sense of self.
5. I strongly desire to have this characteristic.
6. The types of things I do in my spare time (e.g., hobbies) clearly identify me as having this characteristic.

Appendix 2.F Experimental instructions

The following section contains experimental instructions translated from German.

2.F.1 Dictator: *Control*, *Self-image* and *Social image* treatment

The instructions were printed and left in front of subjects' computer screens before they entered the lab. At the beginning of the experiment, the instructions were read out loud by one of the experimenters.

Welcome to this study!

You are participating in an economic study. Depending on your answers, you can earn a certain amount of money. The money will be paid out at the end of the study in cash. It is, therefore, very important that you read the instructions carefully, and that you understand them.

Only for Control treatment. As you can see, there is a video playing on your computer screen. This video will also be played during the study.

Only for Self-image treatment. As you can see, there is a camera installed on the computer screen. The image that the camera is capturing is shown on your computer screen in real time. **Please note: No video streams are saved, and only you and no other person can see your camera video.** At the end of the study we will ask you several short questions about the camera technology and camera settings.

Only for Social image treatment. As you can see, there is a camera installed on the computer screen. You can see another participant of the study. Simultaneously, the participant can also see you. **It is strictly forbidden to communicate in any way with this other participant, e.g, through waving, signs, facial expressions, or similar.** This other participant has the role of observer. Your observer has received his own instructions, in which his task is clearly explained. He has only one task, and that is to observe you and your decisions. Your observer sees your computer screen in real time. That means that all the movements that you do with your mouse, and all the decisions that you take during this study, will be seen by your observer. Please note that there can be short delays in the transmission of the camera video. The transmission of your screen and mouse movements occurs with no delay. **Please note: No video streams are saved, only your observer and no other person can see your camera video.** If you disagree with this, you can finish your participation on the study now. At the end of the study we will ask you several short questions about the camera tech-

nology and camera settings. After the end of the study, your observer will leave the laboratory before you.

All statements made in these instructions are true. This holds generally for all studies conducted at the Bonn Laboratory for Experimental Economic Research, and also for this study.

During the study, communication between participants is forbidden. If you have questions, then please direct them only to us. Please, raise your hand and a member of the experimental team will come to answer privately. Violating this rule leads to exclusion from the study.

The decisions that you make in this study do not only influence how much you will earn, but also, how much another participant will earn. From now on, this participant will be called Participant B.

Participant B is taking part in this study at some other time. You will not find out before or after the study who Participant B is. The same holds true for Participant B. Participant B will not find out who you are.

At the beginning of the study **you will receive 10 euros. Participant B will receive 0 euros.**

Your task is to decide how much you want to send to Participant B. You can send as much as you want. Participant B cannot influence your decision. He will be informed about your options and your decision, and subsequently, he will be paid out in cash.

For example, you can keep 7 euros for yourself, and send 3 euros to Participant B. Or, you can keep 10 euros for yourself and send 0 euros to Participant B. Or, you can keep 5 euros for yourself and send 5 euros to Participant B.

If you have questions, please let us know! Otherwise, we will begin immediately.

2.F.2 Observer

The instructions were printed and left in front of subjects' computer screens before they entered the lab.

READ IMMEDIATELY

Welcome to this study!

Important: It is strictly forbidden to communicate with the participant that you see in the video in any way, e.g., through waving, signs, facial expressions, or similar.

In this study you are participating in a role of observer. You only have one task, and that is to observe another participant of the study. For this task, you will receive a payment of 8 euros in cash at the end of the study.

On your screen you can see a video of another participant and his decision screen in real time. That means that you will observe the decisions that this participant takes during the study. At the same time, this participant can also see you through the camera that is installed on your computer screen. **Please note: No video streams are saved. Except for the participant that you see, no other person can see the video from your camera.**

During the study, communication between participants is forbidden. If you have questions, direct them to us. Raise your hand, and a member of the experimental team will come to answer privately. If you violate this rule, you will be excluded from the study.

The instructions that will be read out loud at the beginning of the study are for the subjects that will take decisions. That means, they are intended for the participant that you are observing. Listen carefully, so that you can understand what is the task of this participant.

Appendix 2.G Additional tables and figures

Table 2.1. OLS regressions estimates of treatment effects

Variables	Dependent variable: amount given					
	Control & *Self image* treat.		*Control* & *Social image* treat.		*Self* & *Social image* treat.	
	(1)	(2)	(3)	(4)	(5)	(6)
Self-image	0.389*	0.369*				
	(0.210)	(0.199)				
Social image			0.956***	0.840***	0.568**	0.490**
			(0.220)	(0.209)	(0.228)	(0.212)
Female (=1)		0.341		0.607**		-0.033
		(0.249)		(0.270)		(0.258)
Age		-0.018		0.026		0.030
		(0.017)		(0.018)		(0.019)
Monthly budget		0.001		0.000		-0.000
		(0.000)		(0.001)		(0.000)
Big5: Openness		0.190*		0.116		0.198
		(0.102)		(0.104)		(0.121)
Big5: Extraversion		-0.453***		-0.172		-0.324***
		(0.107)		(0.114)		(0.118)
Big5: Neuroticism		-0.008		-0.183		0.049
		(0.122)		(0.126)		(0.136)
Big5: Agreeable.		0.585***		0.599***		0.789***
		(0.104)		(0.118)		(0.110)
Big5: Conscient.		-0.092		-0.209*		-0.063
		(0.113)		(0.119)		(0.110)
Constant	1.537***	1.530***	1.537***	0.445	1.926***	1.286***
	(0.142)	(0.444)	(0.142)	(0.439)	(0.154)	(0.472)
Observations	377	376	342	341	343	341
R-squared	0.009	0.127	0.053	0.191	0.018	0.159

Notes: The table presents OLS regressions using the amount given as the dependent variable. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 2.2. OLS regressions estimates of treatment effects by gender

Variables	Dependent variable: amount given											
	Control treatment		*Self-image* treatment		*Social image* treatment		*Control* & *Self image* treat.		*Control* & *Social image* treat.		*Self* & *Social image* treat.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female (=1)	1.02***	1.04***	0.27	0.26	0.65*	0.63*	1.02***	1.03***	1.02***	1.03***	0.27	0.27
	(0.28)	(0.27)	(0.32)	(0.33)	(0.35)	(0.35)	(0.28)	(0.27)	(0.28)	(0.28)	(0.32)	(0.33)
Age		-0.00		-0.01		0.05**		-0.01		0.03*		0.03
		(0.02)		(0.03)		(0.02)		(0.02)		(0.02)		(0.02)
Monthly budget		0.00		0.00		-0.00		0.00		0.00		-0.00
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Self-image							0.83**	0.85**				
							(0.34)	(0.33)				
Self-image *female							-0.75*	-0.77*				
							(0.43)	(0.43)				
Social image									1.16***	1.16***	0.34	0.33
									(0.35)	(0.34)	(0.39)	(0.38)
Social image *female									-0.37	-0.39	0.38	0.37
									(0.45)	(0.44)	(0.47)	(0.48)
Constant	0.94***	0.76	1.76***	1.82**	2.10***	0.99*	0.94***	0.87*	0.94***	0.13	1.76***	1.11**
	(0.21)	(0.58)	(0.26)	(0.70)	(0.28)	(0.58)	(0.21)	(0.45)	(0.21)	(0.42)	(0.26)	(0.51)
Observations	188	188	189	188	154	153	377	376	342	341	343	341
R-squared	0.07	0.07	0.00	0.00	0.02	0.05	0.04	0.04	0.10	0.10	0.03	0.03

Notes: The table presents OLS regressions using the amount given as the dependent variable. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Given that the regressions investigate the effect of gender on the amount given, we do not control for Big 5 personality traits given their known correlations with gender (Schmitt et al., 2008).

Table 2.3. OLS regressions estimates of the effect of observer's gender in *Social image* treatment

Variables	Dependent variable: amount given			
	Social image treatment			
	Male subjects		Female subjects	
	(1)	(2)	(3)	(4)
Female observer (=1)	0.223	0.139	-0.390	-0.244
	(0.545)	(0.571)	(0.416)	(0.449)
Age		0.039		0.065**
		(0.036)		(0.032)
Monthly budget		0.000		-0.000
		(0.001)		(0.001)
Big5: Openness		0.247		0.062
		(0.303)		(0.230)
Big5: Extraversion		0.120		0.081
		(0.283)		(0.266)
Big5: Neuroticism		-0.109		-0.164
		(0.336)		(0.243)
Big5: Agreeableness		1.298***		0.502**
		(0.312)		(0.237)
Big5: Conscientiousness		-0.575		-0.151
		(0.348)		(0.199)
Constant	1.952***	1.252	2.975***	1.292*
	(0.390)	(1.005)	(0.321)	(0.677)
Observations	61	61	93	92
R-squared	0.002	0.382	0.010	0.124

Notes: The table presents OLS regressions using the amount given as the dependent variable. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 2.4. Probit regressions estimates of attrition determinants

Variables	Dependent variable: next week's participation	
	(1)	(2)
Give	0.000	0.004
	(0.013)	(0.013)
Self-image	-0.027	-0.012
	(0.060)	(0.058)
Social image	0.031	0.039
	(0.063)	(0.062)
Female (=1)		-0.001
		(0.061)
Age		0.005
		(0.005)
Monthly budget		-0.000
		(0.000)
Big5: Openness		-0.036
		(0.026)
Big5: Extraversion		-0.006
		(0.030)
Big5: Nuroticism		-0.013
		(0.030)
Big5: Agreeableness		-0.027
		(0.031)
Big5: Conscientiousness		0.017
		(0.029)
Observations	257	255

Notes: The table presents probit regressions using next week's participations as the dependent variable. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The reported coefficients represent average marginal effects evaluated at sample means. Two subjects were excluded from the analysis due to providing identical matching codes, while two subjects provided codes that did not match with any of the first week's codes.

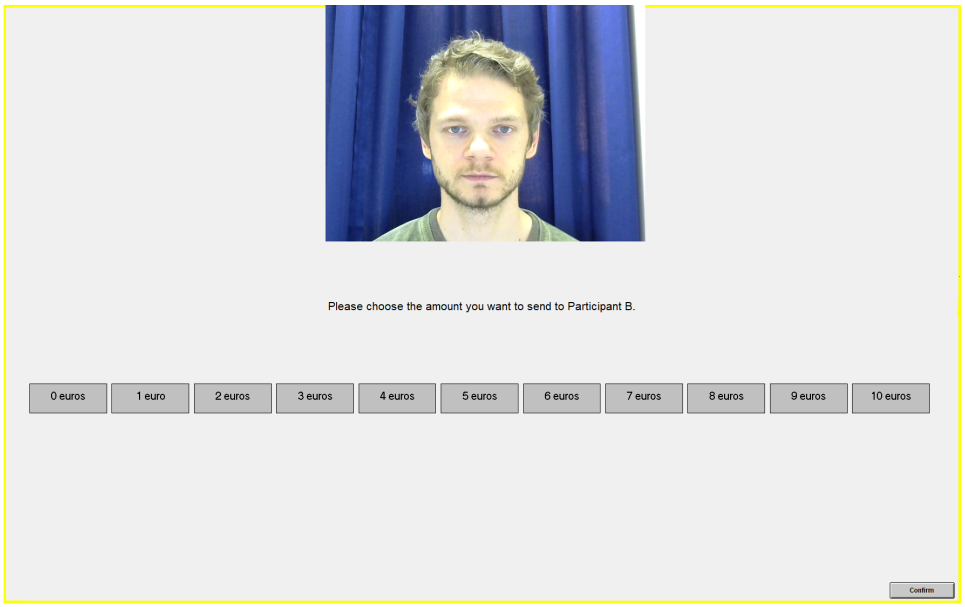

Figure 2.3. Decision screen in *Self-image* and *Social image* treatment

Notes: Translated from German.

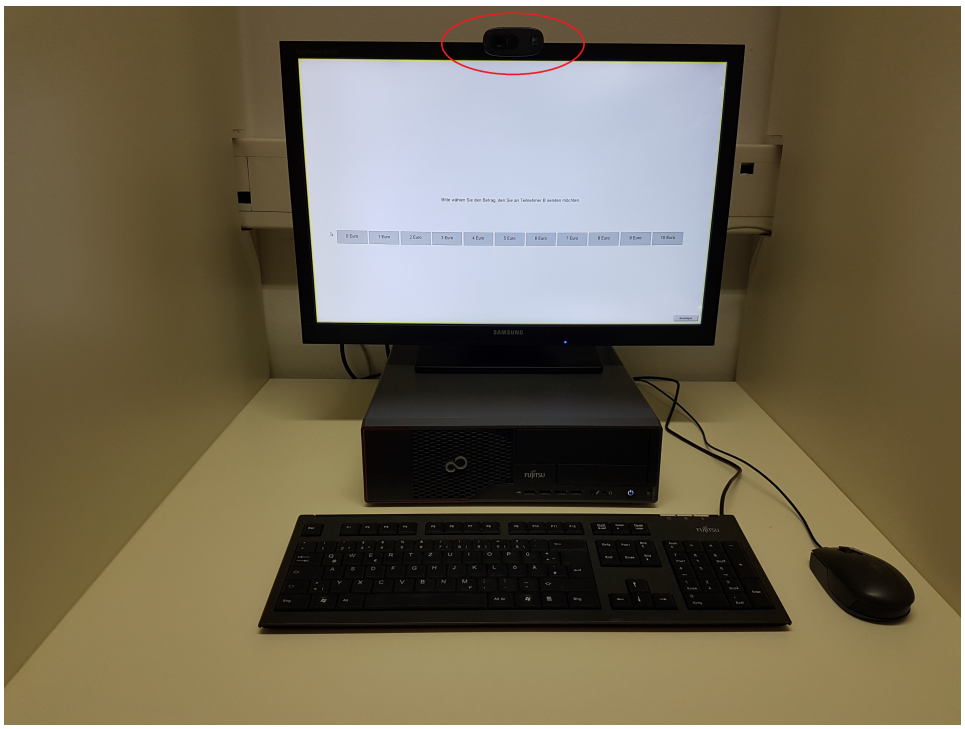

Figure 2.4. The cubicle in *Self-image* and *Social image* treatment with a camera attached to the computer screen

Notes: Video is turned off for demonstration purposes.

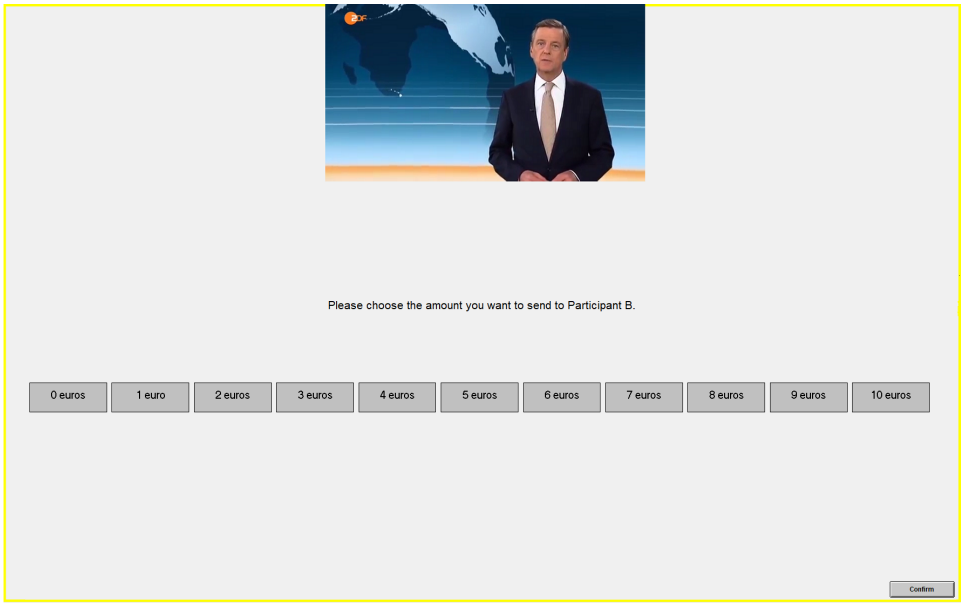

Figure 2.5. Decision screen in *Control* treatment

Notes: Translated from German.

3

The Influence of Self and Social Image Concerns on Lying *

Joint with Armin Falk and Simone Quercia

3.1 Introduction

In recent years, a considerable amount of studies has shown that people experience psychological lying costs as they refrain from lying even when this increases their payoff (e.g. Gneezy, 2005; Mazar et al., 2008; Shalvi et al., 2011a; Fischbacher and Föllmi-Heusi, 2013; Abeler et al., 2014; Gächter and Schulz, 2016). While several explanations have been proposed to characterize lying costs, recent papers show that combining a preference for being honest (intrinsic lying costs) with a preference for being seen as honest (reputation costs) can reconcile the existing empirical findings (Abeler et al., 2016; Gneezy et al., 2018; Khalmetski and Sliwka, 2017).[1] While reputation costs, i.e., social image

* We thank Johannes Abeler, Stefania Bortolotti, Sebastian Kube, Daniele Nosenzo and Matthias Praxmarer as well as workshop and conference participants at Lund University, Maastricht University, University of East Anglia, Humboldt University and Vienna University of Economics and Business for helpful comments. We gratefully acknowledge financial support by Leibniz Programme of the German Science Foundation (DFG) and Institute on Behavior & Inequality (briq).

[1] We use the terminology used in Abeler et al. (2016); Gneezy et al. (2018) use the term direct costs instead of intrinsic costs, while Khalmetski and Sliwka (2017) use the term fixed costs. Also, for the reputation component, Gneezy et al. (2018) use the concept of social identity. While social identity and reputation might not indicate the same constructs in general, in this context they are both used to refer to the willingness to appear honest to external observers. One exception to this modeling approach is M. Dufwenberg and M. A. Dufwenberg (2018), who assume that people suffer only a reputation cost, but differently from the previous models this cost does not depend on the probability of being seen as a liar by an external observer but on the inference that the observer makes on the extent of the lie.

costs arise from the person's desire to appear honest in the eyes of others, the underlying psychological motives for intrinsic costs are debated in the literature. The most prevalent view is that these costs might originate from self-image concerns, that is, the desire to think of oneself as an honest person (Mazar et al., 2008; Shalvi et al., 2011a; Fischbacher and Föllmi-Heusi, 2013).

In this paper, we investigate the influence of these two notions of image concerns on lying behavior using the die-rolling paradigm introduced by Fischbacher and Föllmi-Heusi (2013). In this setup, subjects are given a six-sided die, they are asked to roll it in private and to report the outcome to the experimenter. Payoffs are generally increasing in the report. While lies are not detectable at the individual level, they can be inferred at the group level comparing the distribution of reports with the expected distribution of die rolls.

To make self and social image concerns more salient, we exogenously manipulate self-awareness and observability, which direct the subjects' focus on their private and public self, respectively. To manipulate self-awareness we expose subjects to a real-time video of their face on the computer screen, i.e., we expose them to their "self-image" as in Falk (2017). To manipulate observability, we expose subjects to a real-time video of another subject sitting in the lab while they take their decisions. This other subject also sees the decision maker's face and his or her computer screen in real time, but does not observe his or her die-roll outcome. We compare these two treatments to a *Control* treatment where subjects see a neutral pre-recorded video of another person.

We find that the increase of self-awareness has no significant effect on the average reported die-roll outcome. This suggests that self-image concerns may be less important than previously hypothesized to explain lying behavior, and that intrinsic lying costs might need to incorporate other psychological mechanisms. On the other hand, we show that the increase of observability decreases the average reported outcome even when information about the die-roll outcome is held private. To complement this finding, we conduct a survey experiment where we show that the likelihood of being perceived as a liar increases monotonically with the reported outcome. Moreover, reporting higher outcomes ties subjects with a stigma of likely having undesirable traits in several other dimensions. This further suggests that our effect in the *Social image* treatment indeed stems from the concerns that observers may draw adverse inferences from the observation of high reports.

Our paper contributes to the literature on lying costs in several ways. In particular, numerous studies have suggested self-image or closely related concepts as drivers of intrinsic lying costs. Most closely to our paper, Mazar et al. (2008) investigate a tightly connected notion, one's self-concept. They use religious reminders and honor codes to increase attention to moral standards, and subsequently show that subjects behave more honestly. While using such priming techniques might have its benefits, it has two main drawbacks. First, it primes

everybody towards honest behavior by reminding people of specific moral standards. This, however, does not necessarily imply that these moral standards are the ones congruent with individuals' inner standards. Second, reminding about moral standards might conflate individual with collective standards of behavior.[2] Furthermore, other studies have also hypothesized self-image concerns as determinants of intrinsic lying costs, and have gathered indirect evidence. For example, in their seminal study on the die-rolling paradigm, Fischbacher and Föllmi-Heusi (2013) find that reporting the second highest-paying outcome is perceived as much less dishonest than reporting the highest-paying outcome. The authors suggest that for the subjects that overreport partially, maintaining a favorable self-image might be one of the driving motives. Another example is Shalvi et al. (2011a) who manipulate the number of instructed die rolls, while holding fixed as payoff relevant the first roll. They find that subjects lie more if more die rolls are instructed, and argue that high reports are easier to justify to oneself if observed in any die roll after the first.

In this paper, we design an *exogenous* manipulation where we increase the salience of self-image concerns. In particular, our manipulation builds on the recent work of Bénabou et al. (2018) which implies that self-image concerns arise from the awareness of discrepancy between internal standards of behavior and the self (e.g., in light of current behavior). To make self-image concerns more salient, we manipulate one's self-awareness, increasing the awareness of the aforementioned discrepancy and, ceteris paribus, increasing self-image costs when deviating from internal standards of behavior.[3] This reasoning dates back to the objective self-awareness theory (Duval and Wicklund, 1972), which posits that high levels of self-awareness induce behavior driven by salient moral standards. To test this theory, the most common manipulation has been to place a mirror in front of the subjects during the decision phase. It was used to show that increased self-awareness, for example, decreases simple transgressions (Beaman et al., 1979), increases the attribution of causality for a specific consequence to oneself (Duval and Wicklund, 1973), and induces the usage of corporal punishment depending on the subject's inner attitudes towards it (Carver, 1975).[4]

[2] For this purpose, our design abstracts from reminding subjects of a specific set of morals, but purely emphasizes the salience of inner standards, whatever they might be.

[3] Theoretically this can be related to intrinsic lying costs under perfect information assuming that the motive for this cost is to see oneself as honest (see, e.g., Abeler et al., 2016), or alternatively to imperfect information settings where the agent is assumed to forget his or her "type" and makes inferences about it given the actions taken. This notion is suggested by M. Dufwenberg and M. A. Dufwenberg (2018) as one possible interpretation of their model of lying costs and is in line also with other more general models (e.g. Bénabou and Tirole, 2006, 2011; Bodner and Prelec, 2003).

[4] Note, however, that the mirror manipulation has several potential confounds that we circumvent using a camera manipulation and a specially designed *Control* treatment (see Section 3.2).

Our *Social image* treatment connects to a large body of literature which posits and shows that when being observed, people favor societal standards of behavior (see Ariely et al., 2009; DellaVigna et al., 2012). More specifically, it investigates the effect of increased observability on honesty, which to the best of our knowledge, has been studied so far only in two instances in economics. In particular, Abeler et al. (2016) and Gneezy et al. (2018), in order to test predictions from their theoretical models, design experiments similar to the standard die-rolling paradigm except that subjects see the outcome of a randomizing device on their screens. This allows the experimenter to map the observed outcome to the report for each individual and observe lying behavior at the individual level. Hence, these experiments increase observability by removing private information vis-à-vis the experimenter.[5] While this is reasonable in order to test the above mentioned theories, it comes at a high cost. Specifically, the usual purpose of lying is to deceive a receiver of the information about some private information that the sender has. Thus, to deceive a receiver, some level of uncertainty about the truth on his or her part is required. In contrast to the previous studies, our aim is to investigate the effect of social image concerns in such standard lying situations. Therefore, we maintain private information in our paradigm, but make social image concerns more salient. We do so by keeping the die-roll outcome private, and increasing the observability vis-à-vis the observer, who is informed about the die roller's report and can link the report to his or her identity.

From the perspective of recent theoretical models, we interpret our manipulation of observability as an exogenous increase in the individual parameter governing the reputational payoff, that is, how much subjects care about reputation. This interpretation relies on several plausible observations: i) being exposed to another observer beyond the experimenter may increase the social pressure on the decision maker, ii) the fact that observers can tie the identity of decision makers with their reports may make reputational concerns stronger, and iii) other student participants may constitute a more relevant audience than the experimenter for the decision makers, as this is the audience they would usually be exposed to in everyday life. An increase in the reputation parameter predicts a decrease in lying in models that do not allow downward lying (see Example 4 in M. Dufwenberg and M. A. Dufwenberg (2018) and Proposition 5 in Khalmetski and Sliwka (2017)). Our results confirm this prediction and go in the same direction as the full observability treatments in Abeler et al. (2016) and Gneezy et al. (2018), but we observe a smaller effect than the ones they report (see

[5] Some studies attempt to *reduce* observability vis-à-vis the experimenter by performing double-blind procedures and manipulating probabilities of getting caught; however, they do not report a significant impact on behavior (Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013).

Section 3.4), suggesting that private information has indeed a crucial impact on lying behavior.[6]

The remainder of the paper is structured as follows. In Section 3.2, we describe the experimental design and procedures. In Section 3.3, we report the results of our study. In Section 3.4, we discuss the results and conclude.

3.2 Experimental design and procedure

Our experimental setup is closely based on the die-rolling paradigm introduced by Fischbacher and Föllmi-Heusi (2013). Subjects were asked to roll a six-sided die and report the outcome of the die roll on their computer. Depending on their report they were able to earn any amount from 0 to 5 euros. The payoff is equal to the reported outcome minus 1 euro, i.e., for a report of 1 subjects earned 0 euros, for a report of 2 subjects earned 1 euro, etc. As in Fischbacher and Föllmi-Heusi (2013) subjects were told to roll the die minimally twice, which can facilitate lying; however, they were explicitly told to report the outcome of the first die roll (for experimental instructions see Appendix 3.B). Subjects were asked to roll the die in a non-transparent plastic cup. The cup ensured that only they can observe the outcome which was visible only from directly above the cup.[7]

We designed three different treatments: *Self-image*, *Social image* and *Control* treatment. In the *Self-image* treatment, we exogenously manipulated self-awareness, a mediator of a person's focus on his or her private self. In order to increase self-awareness, we exposed the subjects to their own image. In particular, from the moment subjects sat in the cubicles, a camera installed on the top of the monitor was capturing the image of their face, and playing it in real time on their computer screen (see Figures 3.5 and 3.6 in Appendix). The camera was positioned in a way that the subjects could not evade its visual field, but also, that it was obvious that the plastic cup on the table was outside of this visual field. Additionally, we used a software which automatically detected and zoomed on subjects' faces. The video was placed in the middle-upper part of the screen, while the instructions and the decision screen were placed below it. The subjects were fully informed that the video was not being recorded, and that only they were able to see it. In order to give some meaning to the camera,

[6] In the model by Abeler et al. (2016) that allows for downward lying, an increase in the reputation parameter may have two counterbalancing effects. On the one hand, it makes high reports more costly, and hence induces people to report lower numbers; on the other hand, there will be less expected liars at the high reports and (potentially) more expected liars at the low reports, which would make high reports become more attractive. In this framework, our data would indicate that the first effect dominates the second because people report on average a smaller outcome when they are observed.

[7] The setup is similar as the die-under-cup paradigm by Shalvi et al. (2011b).

subjects were also informed that they would answer a few short questions on the camera technology and settings at the end of the experiment.

In the *Social image* treatment, we exogenously manipulated observability by exposing subjects to the observation of other participants, i.e., observers. Upon their arrival to the laboratory, subjects were randomly assigned one of two roles, decision makers or observers. Decision makers were facing identical procedures and decisions as in *Self-image* treatment, but instead of viewing their face in the video, they saw the face of their observer in real time. Observers were also seated in private cubicles at the same time as the decision makers but had no decisions to make. Each observer was paired with one decision maker. Each observer saw i) the video footage of their paired subject's face in real time and ii) the decision screen of their paired subject in real time. Hence, both the observer and the paired subject saw each other, and additionally, the observer saw the decision maker's screen. This was common knowledge. The procedure made observers fully aware of the reported outcome but not of the actual die outcome. At the end of the decision making part, observers left the laboratory before decision makers. This was publicly announced at the beginning of the study. Alongside the questionnaires from *Self-image* treatment, decision makers were also asked if they have ever seen their observer before, and if so, what was their relationship with them. Only one subject indicated knowing the paired observer. Removing this subject from the data does not change our results.

We compare the decisions in *Self-image* and *Social image* to a *Control* treatment. To design a comparable *Control* treatment we address two concerns. First, the subjects in both of our treatments were exposed to a video. If such a distraction drains cognitive resources, then subjects could be more inclined to act affectively, following their automatic response and potentially biasing the results.[8] Second, in both treatments subjects saw a person who is looking at them: the observer in *Social image* treatment, and themselves in *Self-image* treatment. Several studies have shown that being exposed to simple social cues such as a pair of observing eyes can influence one's behavior (Haley and Fessler, 2005; Bateson et al., 2006; Rigdon et al., 2009).[9] To address these two issues, *Control*

[8] Dishonest behavior has been considered a cognitively demanding process linked to brain areas responsible for cognitive control (Sip et al., 2008; Greene and Paxton, 2009). On the one hand, studies show that when cognitive control is low due to cognitive depletion, people's automatic response is to act more selfishly (Achtziger et al., 2015), and more dishonestly (Gino et al., 2011), which could bias our results upwards. On the other hand, other studies suggest that people's automatic response is to behave more prosocially (Rand et al., 2012; Schulz et al., 2014). If subjects care about how much money the experimenter is left with, or if prosociality as a positive trait is connected to honesty, this could bias our results downwards.

[9] Note, however, that several studies fail to find an effect of social cues in different settings (Fehr and Schneider, 2010; Lamba and Mace, 2010), and question the validity of previous evidence (Carbon and Hesslinger, 2011).

treatment is identical to *Self-image* and *Social image* treatment, except for the following: Instead of seeing their own face or the observer's face, the subjects saw a mute video footage of a famous German news presenter (see Figure 3.7 in Appendix). As the context of the video was immediately recognizable, the subjects were perfectly aware that the video was prerecorded. Moreover, the news presenter is a public person working for a mainstream public-service, and as such does not trigger any tendentious associations.[10]

The lying task lasted less than 10 minutes; hence, it was run right after another experiment, in line with Fischbacher and Föllmi-Heusi (2013). In the experiment preceding the lying task, subjects were participating in another economic game and were exposed to the identical treatment manipulations as in the subsequent die-rolling experiment.[11] This means that the same manipulations were present from the moment subjects entered the cubicles until reaching the short questionnaire at the end of the session. Hence, the attached cameras were not abruptly turned on when reaching the lying task, and their function was clear from the beginning of the session. A total of 685 subjects participated in the study (59.7% female), out of which 531 subjects participated as decision makers and 154 as observers. In *Social image* treatment, the observers earned 8 euros each for their participation in the entire session. Decision makers were not informed about the payment of observers to avoid influences on their behavior driven by social comparison such as, e.g., inequity considerations (Fehr and Schmidt, 1999). Subjects were primarily students of the University of Bonn and were recruited with hroot (Bock et al., 2014). Sessions were conducted at the BonnEconLab. The experiment was programmed using Z-Tree experimental software (Fischbacher, 2007).

In addition to our main experiment, we report results from a short survey which was ran on a separate set of 100 subjects. The survey aims to complement our *Social image* treatment findings by exploring whether people i) perceive that higher reports are more likely to be a product of a lie, and ii) attribute negative traits to subjects who report high numbers. The questions are highly relevant in understanding potential reputation effects which can come from reports. Specifically, given that we maintain private information in our paradigm, it is necessary to see whether reports alone can send undesirable signals about

[10] Comparing *Self* or *Social image* treatment with the *Control* treatment ensures that any difference can be interpreted as the effect of increased self-awareness or reputation, respectively, as any potential effects of social cues or cognitive depletion would be present in all three treatments. Notice, furthermore, that Falk (2017) uses an identical control treatment in an investigation of self-image concerns for moral behavior and compares it with a control treatment without any video, finding no difference between the two.

[11] In the preceding experiment, subjects played a dictator game (see Bašić et al., 2018). We report robustness checks in Appendix 3.A showing that spillover effects cannot account for our findings.

a die-roller. Upon accepting to participate in the survey, subjects were shortly informed about the die-rolling paradigm. Then, they were asked to evaluate the probability of a subject being a liar conditional on each of the six reports. Moreover, for hypothetical reports they were confronted with 6 different statements. For each of the statements they had to indicate how much they agree with it on a scale from 0 to 7. The statements were: "I find this person trustworthy", "I would take this person in a shared flat", "I would lend money to this person", "I would employ this person", "I would buy a car from this person" and "I would vote for this person". The hypothetical reports differed within-subject across the six statements, and the order of the hypothetical reports was randomized between-subject. The survey was conducted with students in front of the University of Bonn library and canteen. Each participant earned 5 euros for participation.

3.3 Results

We divide our results section in two subsections. In the first, we report the comparison between the *Control* and the *Self-image* treatment. In the second, we contrast the results of the *Control* and the *Social image* treatment, and complement this comparison with the results of our survey experiment.

3.3.1 The influence of self-image

In this subsection, we focus on the difference between the *Control* ($n = 188$) and the *Self-image* treatment ($n = 189$). First, we check if people overreport the outcome of die rolling. To do that, we contrast the reports of the die rolls with the uniform distribution by applying the Kolmogorov-Smirnov test for discrete data (henceforth KS d; Jann, 2008). In Figure 3.1 we report the frequency of each reported outcome (left panel) and the average reported outcome (right panel) in the two treatments. We find that the distribution of reports is significantly different from the uniform distribution in both treatments (two-sided KS d test, $p < 0.001$ for both treatments). Next, we compare average reports between the two treatments. Figure 3.1 (right panel) shows that people report on average 4.62 in *Self-image* treatment and 4.70 in *Control* treatment with no significant differences across the two (two-sided t-test, $p = 0.630$). The result remains insignificant if we perform the analysis using an OLS regression and including control variables (see Table 3.1 in the following subsection). Next, we compare the distributions between the two treatments (Figure 3.1, left panel). Also in this case, we do not find any significant difference (Fisher's exact test, $p = 0.310$; Epps-Singleton two-sample test, $p = 0.270$).

Finally, we turn to the analysis of report frequencies for each possible outcome. We observe that in both *Control* and *Self-image* treatment people overreport the outcome 5 (two-sided binomial test against the expected true value

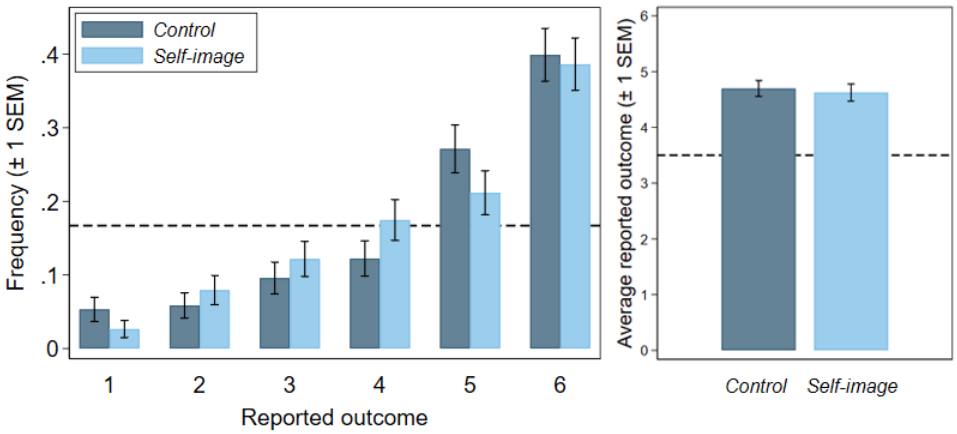

Figure 3.1. Frequencies of each reported outcome (left panel) and the average reported outcome (right panel) in *Self-image* and *Control* treatment

Notes: Error bars indicate standard error of the means. The dashed line represents the expected frequency of each outcome (left panel) and the expected average outcome (right panel).

of 0.167, $p < 0.001$ for *Control* treatment, $p = 0.097$ for *Self-image* treatment) and the outcome 6 (two-sided binomial test against the expected true value of 0.167, $p < 0.001$ for both treatments). If we compare the two distributions, we observe that the frequencies of all outcomes are very similar across the two treatments. Only the frequencies for outcomes 4 and 5 exhibit a noticeable difference across treatments; however, this difference is not statistically significant (two-sided binomial test of proportions, $p = 0.154$ for outcome 4, $p = 0.183$ for outcome 5).

Overall, higher self-awareness and emphasized self-image concerns induce neither a significant difference in the average reported outcome nor a significant difference in the distributions of the reports or the report frequencies of any specific outcome.

3.3.2 The influence of social image

In this subsection, we analyze the difference between the *Control* ($n = 188$) and the *Social image* treatment ($n = 154$). In Figure 3.2 we report the frequency of each reported outcome (left panel) and the average reported outcome (right panel) in the two treatments. First of all and similar to *Control* and *Self-image* treatment, subjects in *Social image* treatment significantly overreport their die-roll outcome compared to the uniform distribution (two-sided KS d test, $p < 0.001$). However, as seen in Figure 3.2 (right panel), people on aver-

age report less in *Social image* (4.34) than they do in *Control* treatment (4.70). This difference is statistically significant (two-sided t-test, $p = 0.038$). When comparing observed distributions (Figure 3.2, left panel), we do not find a significant difference (Fisher's exact test, $p = 0.351$; Epps-Singleton two-sample test, $p = 0.344$).

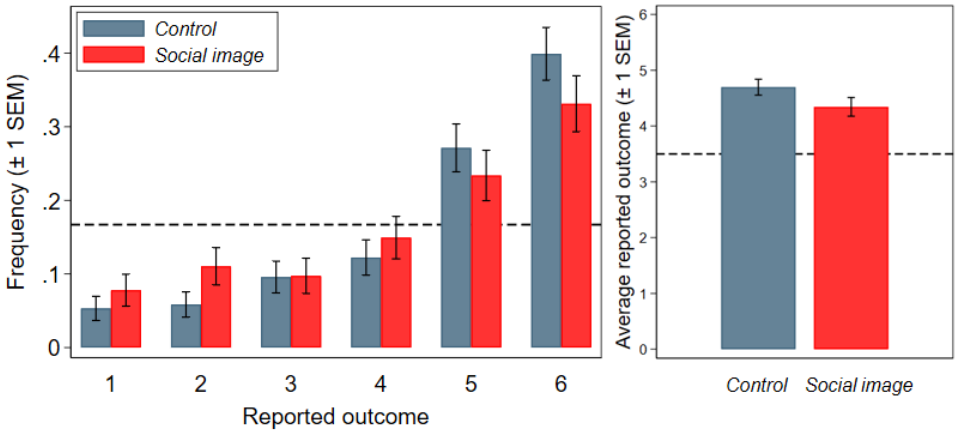

Figure 3.2. Frequencies of each reported outcome (left panel) and the average reported outcome (right panel) in *Social image* and *Control* treatment

Notes: Error bars indicate standard error of the means. The dashed line represents the expected frequency of each outcome (left panel) and the expected average outcome (right panel).

Next, we focus on the reports for each possible outcome. Similarly to *Control* treatment, we observe that subjects in *Social image* treatment overreport the outcome 5 and the outcome 6 (two-sided binomial test against the expected true value of 0.167, $p = 0.025$ for outcome 5, $p < 0.001$ for outcome 6). When comparing all 6 possible outcomes across the two treatments, we observe that the percentage of all reported outcomes is closer to the expected true value in *Social image* than in *Control* treatment, which can explain the observed shift of the average reported outcome. However, these changes in report frequencies are rather small, and most outcomes in *Social image* treatment are not significantly different from *Control* treatment (two-sided binomial test of two proportions, $p > 0.196$ for outcomes 1, 3, 4, 5 and 6, $p = 0.082$ for the outcome 2).

Next, in Table 3.1 we report an OLS regression analysis to confirm the robustness of our findings in *Self-image* and *Social image* treatment to further controls.[12] We report 3 regression models where we use the reported outcome

[12] The results stay very similar when using ordered probit estimations.

Table 3.1. OLS regressions estimates of treatment effects

Variables	Dependent variable: reported outcome		
	(1)	(2)	(3)
Self-image	-0.072	-0.069	-0.054
	(0.150)	(0.148)	(0.149)
Social image	-0.353**	-0.346**	-0.309*
	(0.171)	(0.169)	(0.168)
Female (=1)		-0.516***	-0.371**
		(0.130)	(0.150)
Age			-0.002
			(0.015)
Monthly budget			0.000
			(0.000)
Big5: Extraversion			0.216***
			(0.075)
Big5: Agreeableness			-0.257***
			(0.072)
Big5: Neuroticism			0.039
			(0.081)
Big5: Conscientiousness			-0.048
			(0.071)
Big5: Openness			0.026
			(0.069)
Constant	4.697***	5.001***	4.875***
	(0.108)	(0.128)	(0.364)
Observations	531	531	529
R-squared	0.009	0.037	0.069

Notes: The table presents OLS regressions using reported outcome as the dependent variable. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

as dependent variable and two treatment dummies (with *Control* treatment as omitted category). Model (1) reports these estimates without controls and confirms the results from non-parametric tests. Model (2) indicates that these findings are robust when controlling for gender. Model (3) shows the same results also controlling for age, monthly budget and personality characteristics (*Social image* treatment remains significant at a 10% level; $p = 0.067$). In addition, we also show that females report less than males in the overall sample replicating a standard finding in the die-roll experiments (see Abeler et al., 2016). Fur-

thermore, we observe that reports increase with extraversion and decrease with agreeableness.

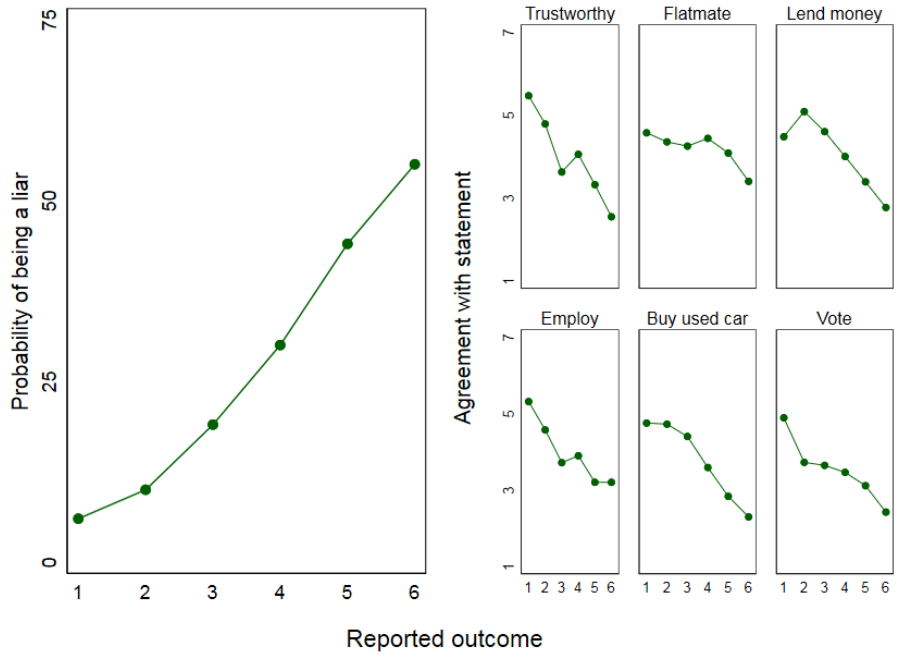

Figure 3.3. Survey results

Notes: The average of the expected probability that the subject is lying conditioning on the reported outcome (left panel), and the average agreement with the statement conditioning on the reported outcome (right panel). The statements are "I find this person trustworthy", "I would take this person in a shared flat", "I would lend money to this person", "I would employ this person", "I would buy a car from this person" and "I would vote for this person".

Finally, we report the results of the survey conducted on uninvolved subjects ($n = 100$) to complement our social image results. In Figure 3.3 we report the average of the expected probability that a subject is a liar conditional on the report (left panel), and the average agreement with the statements conditional on the subject's report (right panel). The average belief about the probability that subject is a liar increases with the report (OLS regression with standard errors clustered at the individual level, $p < 0.001$). In particular, we observe a monotonic increase of beliefs ranging from 6% to 55%. With respect to the agreement with the statements, we observe that on average, survey participants perceive a person reporting higher numbers as less trustworthy, and they declare to be less willing to consider such a person as a flatmate, to lend them money,

employ them, buy a car from them or vote for them (Spearman's rho, $p = 0.004$ for "Flatmate" comparison, $p < 0.001$ for all other comparisons).

To sum up, we observe that increased observability decreases the average reported outcome. Furthermore, we observe that reporting higher numbers signals higher likelihood of being a liar, and higher likelihood of being untrustworthy and having undesirable traits in many other domains. This further supports the reputation channel as the explanation behind the results of our social image manipulation. Specifically, on average, the observer who sees a high report in contrast to a low report will think that the die roller is more likely to be a liar and more likely to have undesirable traits. This, in turn, gives rise to reputation costs for the die roller, and can explain why the die roller can benefit from reporting a lower number.

3.4 Discussion and conclusion

In this paper we have investigated the influence of self and social image concerns on lying behavior. We have exogenously manipulated self-awareness and observability by exposing subjects to their own image or to the observation of another participant in real time, respectively. We have shown that the increase of self-awareness through our manipulation has no effect on the average report.

There are several reasons why this may be the case. One possibility is that the importance of self-image concerns might be very low in certain domains. Falk (2017) uses a similar self-image manipulation as ours, and shows that subjects care about their self-image when confronted with the choice of administering an electric shock to another individual for money. Moreover, using an identical manipulation, Bašić et al. (2018) find that increasing self-awareness significantly increases generosity. Hence, it could potentially be that honesty is not a (salient) inner standard, and the act of lying has a negligible impact on one's self image. Alternatively, self-image concerns could be an important determinant of lying and cheating behaviors; but in the die-rolling task, where the only "victim" of immoral behavior is the experimenter, and the negative externality is arguably weaker than when administering an electric shock or being selfish to another participant, the strength of self-image concerns could be insufficient to generate a significant shift in behavior. Hence, it would be important to study self-image concerns in settings where the externality of lies is more pronounced as for example sender-receiver games (Crawford and Sobel, 1982; Gneezy, 2005).[13]

[13] Additionally, it is possible that for a certain proportion of subjects, their inner standard is self-interest and not honesty, which could in turn explain why the average report does not change in our self-image manipulation. This explanation, however, is not consistent with standard conceptualizations of self-image concerns (e.g., Bénabou and Tirole, 2006; Bénabou et al., 2018). Moreover, with such polarization of inner standards, we would also expect a polarization of re-

The lack of self-image effect might suggest that there could be other psychological mechanisms that lie at the origin of intrinsic lying costs in die-rolling experiments.[14] While several motives have already been proposed, e.g., social norms and guilt aversion, it was recently shown that these motives cannot reconcile all the findings from previous die-rolling experiments (Abeler et al., 2016). One potential explanation is that honesty could have components of heuristical behavior, that is, subjects display automatic honest behaviors that have been instilled by parents and/or other authority figures (see Bénabou et al., 2018), and do not question these behaviors even if characteristics of the decision environment change. Alternatively, in line with M. Dufwenberg and M. A. Dufwenberg (2018) subjects could have no intrinsic lying costs but reputational concerns vis-à-vis an external audience with respect to the size of their lies. M. Dufwenberg and M. A. Dufwenberg (2018) show that their model can generate a distribution of reports in line with a standard die-rolling experiment in the case where the only audience is the experimenter.

With respect to social image concerns, we have shown that the increase in subjects' observability significantly decreases the average report, and that reporting high paying numbers ties subjects with a stigma of likely being a liar and having undesirable traits. The results cleanly indicate that reputation constitutes an important component of lying costs. Different from previous literature, however, we have reported a manipulation that keeps information about the true outcome of the die roll private, and hence maintains the private-information property of standard lying situations. The direction of our effect confirms theories that predict a decrease in lying due to increased reputational concerns (M. Dufwenberg and M. A. Dufwenberg, 2018; Khalmetski and Sliwka, 2017). As we have noted, we find a smaller effect compared to the effects found when the outcome of a randomizing device is public. In particular, we observe a Cohen's d effect size of 0.226, while Abeler et al. (2016) observe an effect size of 0.761 and Gneezy et al. (2018) of 0.284. This suggests that private information has indeed an important impact on lying behavior.

---

ports in the *Self-image* treatment, which is not supported by our data. Finally, it is also possible that subjects' self-awareness in the die-rolling paradigm is rather high in the *Control* treatment; hence, manipulating it cannot provoke a sufficient further increase. This, however, is at odds with studies that also use the camera manipulation and find an effect of self-awareness using different games, which would imply that the level of self-awareness would need to be highly context specific: higher in the die-rolling paradigm than in a morality paradigm (Falk, 2017) or a dictator game (Bašić et al., 2018).

[14] Note that the existence of intrinsic lying costs was identified in the die-rolling paradigm (Abeler et al., 2016; Gneezy et al., 2018; Khalmetski and Sliwka, 2017).

References

Abeler, Johannes, Anke Becker, and Armin Falk (2014): "Representative evidence on lying costs." *Journal of Public Economics*, 113, 96–104. [33]

Abeler, Johannes, Daniele Nonsenzo, and Collin Raymond (2016): "Preferences for truth-telling." *CESifo Working Paper* (6087). [33, 35–37, 43, 46]

Achtziger, Anja, Carlos Alós-Ferrer, and Alexander K. Wagner (2015): "Money, depletion, and prosociality in the dictator game." *Journal of Neuroscience, Psychology, and Economics*, 8 (1), 1–14. [38]

Ariely, Dan, Anat Bracha, and Stephan Meier (2009): "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *The American Economic Review*, 99 (1), 544–555. [36]

Bašić, Zvonimir, Armin Falk, and Simone Quercia (2018): "Self-image, social image and prosocial behavior." [39, 45, 46, 50]

Bateson, Melissa, Daniel Nettle, and Gilbert Roberts (2006): "Cues of being watched enhance cooperation in a real-world setting." *Biology Letters*, 2 (3), 412–414. [38]

Beaman, Arthur L., Bonnel Klentz, Edward Diener, and Soren Svanum (1979): "Self-awareness and transgression in children: Two field studies." *Journal of Personality and Social Psychology*, 37 (10), 1835–1846. [35]

Bénabou, Roland, Armin Falk, and Jean Tirole (2018): "Narratives, imperatives and moral reasoning." *Working Paper*. [35, 45, 46]

Bénabou, Roland and Jean Tirole (2006): "Incentives and prosocial behavior." *The American Economic Review*, 96 (5), 1652–1678. [35, 45]

Bénabou, Roland and Jean Tirole (2011): "Identity, morals, and taboos: Beliefs as assets." *The Quarterly Journal of Economics*, 126 (2), 805–855. [35]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014): "hroot: Hamburg registration and organization online tool." *European Economic Review*, 71, 117–120. [39]

Bodner, Ronit and Drazen Prelec (2003): "Self-signaling and diagnostic utility in everyday decision making." In *The Psychology of Economic Decisions*. Ed. by Isabelle Brocas and Juan D Carillo. Oxford University Press, 105–126. [35]

Carbon, Claus-Christian and Vera M. Hesslinger (2011): "Bateson et al.'s (2006) cues-of-being-watched paradigm revisited." *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*, 70 (4), 203–210. [38]

Carver, Charles S. (1975): "Physical aggression as a function of objective self-awareness and attitudes toward punishment." *Journal of Experimental Social Psychology*, 11 (6), 510–519. [35]

Crawford, Vincent P. and Joel Sobel (1982): "Strategic Information Transmission." *Econometrica*, 50 (6), 1431–1451. [45]

DellaVigna, Stefano, John A. List, and Ulrike Malmendier (2012): "Testing for altruism and social pressure in charitable giving." *The Quarterly Journal of Economics*, 127 (1), 1–56. [36]

Dufwenberg, Martin and Martin A. Dufwenberg (2018): "Lies in disguise – A theoretical analysis of cheating." *Journal of Economic Theory*, 175, 248–264. [33, 35, 36, 46]

Duval, Shelley and Robert A. Wicklund (1972): *A theory of objective self awareness*. Oxford: Academic Press. [35]

Duval, Shelley and Robert A. Wicklund (1973): "Effects of objective self-awareness on attribution of causality." *Journal of Experimental Social Psychology*, 9 (1), 17–31. [35]

Engel, Christoph (2011): "Dictator games: A meta study." *Experimental Economics*, 14 (4), 583–610. [50]

Falk, Armin (2017): "Facing yourself – A note on self-image." *CESifo Working Paper Series* (6428). [34, 39, 45, 46]

Fehr, Ernst and Klaus M. Schmidt (1999): "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114 (3), 817–868. [39]

Fehr, Ernst and Frédéric Schneider (2010): "Eyes are on us, but nobody cares: Are eye cues relevant for strong reciprocity?" *Proceedings of the Royal Society of London B: Biological Sciences*, 277 (1686), 1315–1323. [38]

Fischbacher, Urs (2007): "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10 (2), 171–178. [39]

Fischbacher, Urs and Franziska Föllmi-Heusi (2013): "Lies in disguise — an experimental study on cheating." *Journal of the European Economic Association*, 11 (3), 525–547. [33–37, 39, 50]

Gächter, Simon and Jonathan F. Schulz (2016): "Intrinsic honesty and the prevalence of rule violations across societies." *Nature*, 531 (7595), 496–499. [33]

Gino, Francesca, Maurice E. Schweitzer, Nicole L. Mead, and Dan Ariely (2011): "Unable to resist temptation: How self-control depletion promotes unethical behavior." *Organizational Behavior and Human Decision Processes*, 115 (2), 191–203. [38]

Gneezy, Uri (2005): "Deception: The role of consequences." *The American Economic Review*, 95 (1), 384–394. [33, 45]

Gneezy, Uri, Agne Kajackaite, and Joel Sobel (2018): "Lying aversion and the size of the lie." *American Economic Review*, 108 (2), 419–53. [33, 36, 46]

Greene, Joshua D. and Joseph M. Paxton (2009): "Patterns of neural activity associated with honest and dishonest moral decisions." *Proceedings of the National Academy of Sciences*, 106 (30), 12506–12511. [38]

Haley, Kevin J. and Daniel M. T. Fessler (2005): "Nobody's watching?: Subtle cues affect generosity in an anonymous economic game." *Evolution and Human behavior*, 26 (3), 245–256. [38]

Jann, Ben (2008): "Multinomial goodness-of-fit: Large-sample tests with survey design correction and exact tests for small samples." *Stata Journal*, 8 (2), 147. [40]

Khalmetski, Kiryl and Dirk Sliwka (2017): "Disguising lies-Image concerns and partial lying in cheating games." *CESifo Working Paper* (6347). [33, 36, 46]

Khan, Uzma and Ravi Dhar (2006): "Licensing effect in consumer choice." *Journal of Marketing Research*, 43 (2), 259–266. [50]

Lamba, Shakti and Ruth Mace (2010): "People recognise when they are really anonymous in an economic game." *Evolution and Human Behavior*, 31 (4), 271–278. [38]

Mazar, Nina, On Amir, and Dan Ariely (2008): "The dishonesty of honest people: A theory of self-concept maintenance." *Journal of Marketing Research*, 45 (6), 633–644. [33, 34, 36]

Mazar, Nina and Chen-Bo Zhong (2010): "Do green products make us better people?" *Psychological Science*, 21 (4), 494–498. [50]

Rand, David G., Joshua D. Greene, and Martin A. Nowak (2012): "Spontaneous giving and calculated greed." *Nature*, 489, 427–430. [38]

Rigdon, Mary, Keiko Ishii, Motoki Watabe, and Shinobu Kitayama (2009): "Minimal social cues in the dictator game." *Journal of Economic Psychology*, 30 (3), 358–367. [38]

Schulz, Jonathan F., Urs Fischbacher, Christian Thöni, and Verena Utikal (2014): "Affect and fairness: Dictator games under cognitive load." *Journal of Economic Psychology*, 41, 77–87. [38]

Shalvi, Shaul, Jason Dana, Michel J. J. Handgraaf, and Carsten K. W. De Dreu (2011a): "Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior." *Organizational Behavior and Human Decision Processes*, 115 (2), 181–190. [33–35]

Shalvi, Shaul, Michel J. J. Handgraaf, and Carsten K. W. De Dreu (2011b): "Ethical manoeuvring: Why people avoid both major and minor lies." *British Journal of Management*, 22 (s1), S16–S27. [37]

Sip, Kamila E., Andreas Roepstorff, William McGregor, and Chris D. Frith (2008): "Detecting deception: The scope and limits." *Trends in Cognitive Sciences*, 12 (2), 48–53. [38]

Appendix 3.A Robustness check

In line with Fischbacher and Föllmi-Heusi (2013), due to the brevity of the lying task we conducted it after another experiment. In this other experiment subjects played a dictator game. Given that they were exposed to the same manipulation in the dictator game and the lying game, this raises concerns for potential asymmetric spillover effects. In particular, both self and social image manipulations increased generous behavior in the dictator game (see Bašić et al., 2018); hence, this could have encouraged self-licensing, i.e., a subject acts generously in the dictator game, but then this behavior licenses the subject to subsequently lie in the lying task (see Khan and Dhar, 2006; Mazar and Zhong, 2010).[15] We test for this potential concern in two steps. First, we check if in any of the three treatments there is a positive correlation between the dictator giving and the report in the subsequent lying game. Such a result would indicate strong evidence for licensing. We do not find such an effect in any of the three treatments (Spearman's rho < -0.21, $p < 0.004$ for all three treatments). The negative correlation we observe indicates that across all three treatments, on average, subjects that give more in the dictator game also report less in the lying game, suggesting a stable preference for acting in a moral manner across the two games. Second, we test if increased self-awareness or observability causes different lying behavior across subjects that exhibit similar generosity in the dictator game. In particular, it could be that even if subjects are acting in a moral manner across games, they engage in subtle licensing, where due to more generous behavior in the two image treatments, subjects are encouraged to report slightly more than subjects that exhibit similar generosity in the *Control* treatment. To investigate this on a detailed level, we divide our sample in categories according to their behavior in the dictator game. We observe that the majority of our subjects (98.49%) fit into one of the three categories: 42.37% of subjects give 0% of their endowment, 39.55% give between 0% and 50%, and 16.57% of subjects give 50%.[16] We test the difference in reports from the lying task for each of the three categories across *Control* and *Self-image* treatment, and *Control* and *Social image* treatment and find no significant effects (two-sided t-test, $p > 0.227$ for all six comparisons). Hence, altogether, we do not find any evidence that spillovers from previous experiment can account for the differences in lying behavior, or lack thereof, across treatments.

[15] Note that we observe an effect in the social image manipulation; hence, regarding social image, moral licensing could only cause an underestimation of the effect size as it would motivate subjects to lie more.

[16] This distribution of behavior is in line with other studies (see Engel, 2011).

Appendix 3.B Experimental instructions

The following section contains experimental instructions translated from German.

3.B.1 Dictator: *Control*, *Self-image* and *Social image* treatment

The general instructions and the instructions for the first game were printed and left in front of subjects' computer screens before they entered the lab. At the beginning of the experiment, the instructions were read out loud by one of the experimenters.

Welcome to this study!

You are participating in an economic study. Depending on your answers, you can earn a certain amount of money. The money will be paid out at the end of the study in cash. It is, therefore, very important that you read the instructions carefully, and that you understand them.

Only for Control treatment. As you can see, there is a video playing on your computer screen. This video will also be played during the study.

Only for Self-image treatment. As you can see, there is a camera installed on the computer screen. The image that the camera is capturing is shown on your computer screen in real time. **Please note: No video streams are saved, and only you and no other person can see your camera video.** At the end of the study we will ask you several short questions about the camera technology and camera settings.

Only for Social image treatment. As you can see, there is a camera installed on the computer screen. You can see another participant of the study. Simultaneously, the participant can also see you. **It is strictly forbidden to communicate in any way with this other participant, e.g, through waving, signs, facial expressions, or similar.** This other participant has the role of observer. Your observer has received his own instructions, in which his task is clearly explained. He has only one task, and that is to observe you and your decisions. Your observer sees your computer screen in real time. That means that all the movements that you do with your mouse, and all the decisions that you take during this study, will be seen by your observer. Please note that there can be short delays in the transmission of the camera video. The transmission of your screen and mouse movements occurs with no delay. **Please note: No video streams are saved, only your observer and no other person can see your camera video.** If you disagree with this, you can finish your participation on the study now. At the end of the study we will ask you several short questions about the camera tech-

nology and camera settings. After the end of the study, your observer will leave the laboratory before you.

All statements made in these instructions are true. This holds generally for all studies conducted at the Bonn Laboratory for Experimental Economic Research, and also for this study.

During the study, communication between participants is forbidden. If you have questions, then please direct them only to us. Please, raise your hand and a member of the experimental team will come to answer privately. Violating this rule leads to exclusion from the study.

At this point, the experimenter read the instructions of the first game, which was followed by the game itself. After the game, we presented the subjects with the instructions of the die-rolling experiment on their computer screens.

Die roll, introductory screen. The first part of the study is now finished. The second part of the study is not connected to the first. For the following task you will require a cup and a die. Please wait, until we bring it to you.

At this point, the subjects were each given a plastic non-transparent cup and a die.

Die roll, instructions screen. Please, do not use the die nor the cup before you are asked to do so. When you will be asked to roll the die, roll it twice. More specifically, take it with your hand and roll it in the cup. Your task it to report which number you rolled with the first try. Depending on the reported number, you will receive a certain amount of money. How much money you earn for a given number is presented in the table underneath *(see Figure 3.4)*.

Rolled number	1	2	3	4	5	6
Payoff	0 euros	1 euro	2 euros	3 euros	4 euros	5 euros

Figure 3.4. Payment table

The second roll is to assure yourself that the die is not loaded. You can roll the die also more than twice; however, only the first roll counts. The money that you earn in this task will be added to the money you earned so far and paid out at the end of the study in cash.

One more time: When you are asked to roll the die, roll it twice. Report the number that you rolled with the first try. The amount of money you will earn depending on your report is presented in the table.

Please, do not start yet. If you understood everything, press "next". If you have questions, raise your hand.

Die roll, decision screen. Roll the die twice now. Report which number you rolled with the first try.

3.B.2 Observer

The instructions were printed and left in front of subjects' computer screens before they entered the lab.

READ IMMEDIATELY

Welcome to this study!

Important: It is strictly forbidden to communicate with the participant that you see in the video in any way, e.g., through waving, signs, facial expressions, or similar.

In this study you are participating in a role of observer. Your only have one task, and that is to observe another participant of the study. For this task, you will receive a payment of 8 euros in cash at the end of the study.

On your screen you can see a video of another participant and his decision screen in real time. That means that you will observe the decisions that this participant takes during the study. At the same time, this participant can also see you through the camera that is installed on your computer screen. **Please note: No video streams are saved. Except for the participant that you see, no other person can see the video from your camera.**

During the study, communication between participants is forbidden. If you have questions, direct them to us. Raise your hand, and a member of the experimental team will come to answer privately. If you violate this rule, you will be excluded from the study.

The instructions that will be read out loud at the beginning of the study are for the subjects that will take decisions. That means, they are intended for the participant that you are observing. Listen carefully, so that you can understand what is the task of this participant.

Appendix 3.C Additional figures

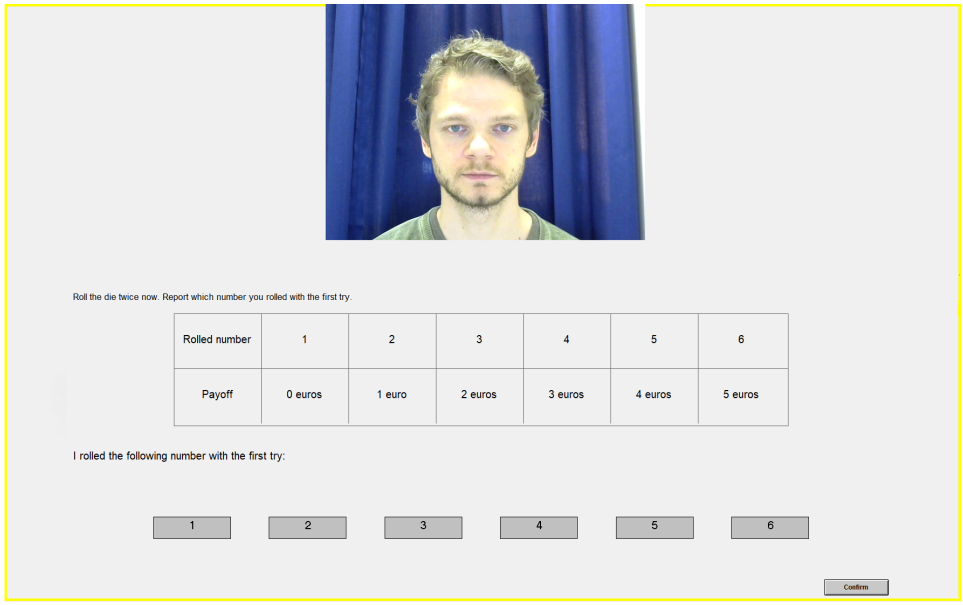

Figure 3.5. Decision screen in *Self-image* and *Social image* treatment

Notes: Translated from German.

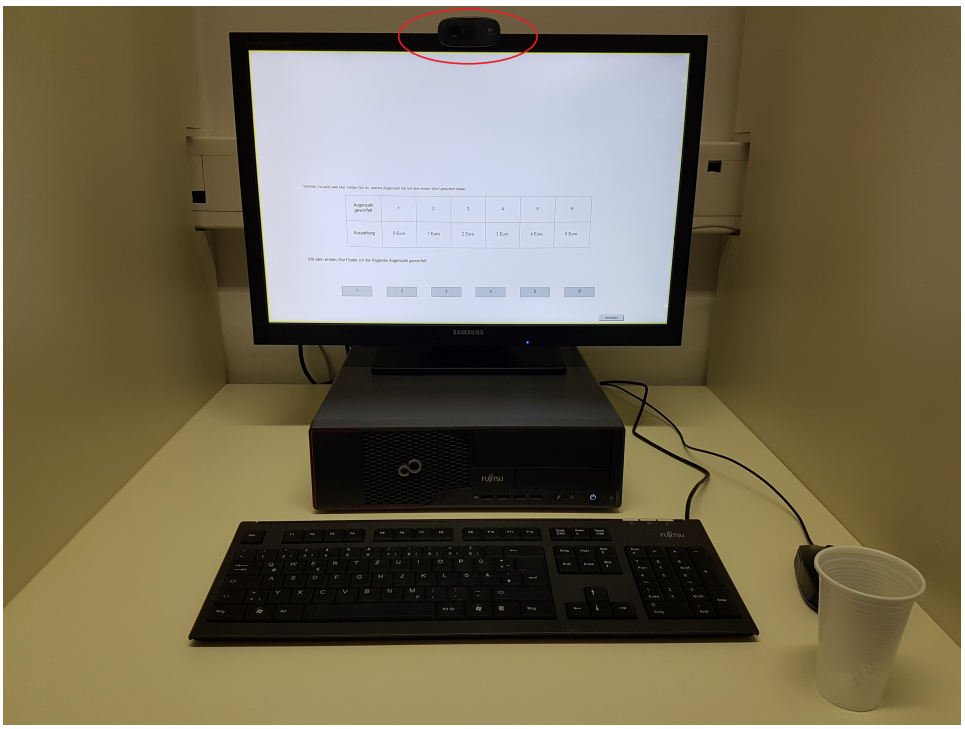

Figure 3.6. The cubicle in *Self-image* and *Social image* treatment with a camera attached to the computer screen

Notes: The video is turned off for demonstration purposes.

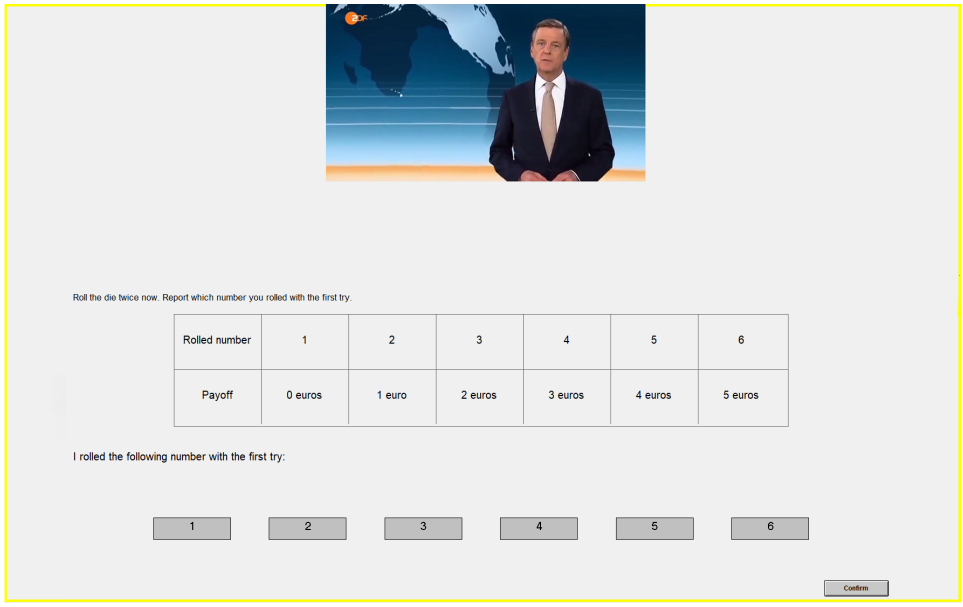

Figure 3.7. Decision screen in *Control* treatment

Notes: Translated from German.

4

The Influence of Self and Social Image Concerns in Childhood and Adolescence *

Joint with Armin Falk and Simone Quercia

4.1 Introduction

Prosocial behavior is ubiquitous in nature. It is critical for a variety of societal functions and outcomes, ranging from charitable giving and volunteering to trade and economic growth (Zak and Knack, 2001; Guiso et al., 2009; Carpenter and Myers, 2010; DellaVigna et al., 2012). Given the importance of prosocial behavior, a large literature is dedicated to understanding its underlying drivers. Recently, a significant amount of both theoretical and empirical work shifted focus towards image concerns. Image concerns refer to an individual's desire to create and uphold a positive image of him- or herself; *self-image concerns* refer to the desire to uphold a positive image in the eyes of oneself, while *social image concerns*, i.e., reputation concerns, refer to the desire to uphold a positive image in the eyes of others. Both were shown to causally influence prosocial behavior (Andreoni and Petrie, 2004; Ariely et al., 2009; Grossman and van der Weele, 2017; Falk, 2017), and are considered as its critical components. While lots of studies focus on the effect of self and social image concerns in adults, very little is known about their roots in young age. This is of particular interest, as i)

* We thank the headmasters of schools and the parents of children for allowing us to perform the research as well as Ana Milić, Antonija Nikolaš, Nikolina Oršićek, Sanja Petrić and Arina Veseli for excellent research support. We also gratefully acknowledge financial support by the Institute on Behavior and Inequality (briq).

it can shed light on the origins of self and social image concerns, ii) it can uncover the roots of the commonly observed gender asymmetry in image concerns, which supports evolutionary theories suggesting that men are more inclined to costly signal prosocial traits (see Trivers, 1972; Amotz Zahavi and Avishag Zahavi, 1999; E. A. Smith and Bird, 2000; Raihani and S. Smith, 2015), and iii) it can illuminate the role of image concerns as an underpinning mechanism of prosocial behavior in children and adolescents.

In this study we investigate the influence of self and social image concerns on prosocial behavior in 7-14 year-old children and adolescents. We do so by implementing a dictator game with 494 subjects, where we exogenously manipulate self-awareness and observability in order to direct the subjects' focus on their private and public self, respectively. We show that both self and social image concerns matter, however, only for boys. More specifically, we show that: i) boys increase their giving with the increase of social image concerns, significantly more so than girls (who do not react), ii) boys increase their giving with the increase of self-image concerns, while girls do not, and iii) the opinion about fairest behavior, together with the perceived fairness of a selfish behavior does not vary across genders, suggesting that the behavioral gender asymmetry is not due to differences in understanding of normative behavior, but due to an actual difference in preferences for being viewed as prosocial. Our results support recent predictions suggesting that reputation should start playing a role at the age of 7-10 (Warneken, 2018). Interestingly, the image effects we observe diminish with older children, which is driven by an increase of giving in the *Control* treatment but a lack of increase in the two image treatments, over age. Specifically, giving in older children comes close to what they think is fair behavior, suggesting that image concerns might be able to increase prosocial behavior as long as inner prosocial preferences are sufficiently weak and leave scope for an increase. Finally, the stark gender findings we observe support evolutionary theories implying that men have a higher benefit of signaling prosocial traits (see Trivers, 1972; Amotz Zahavi and Avishag Zahavi, 1999; E. A. Smith and Bird, 2000; Raihani and S. Smith, 2015), and show that this important gender difference emerges very early in human development.

Our results connect and contribute to several streams of literature. First and foremost, they connect to the literature on image concerns and prosocial behavior in adults. The theoretical models of image concerns propose that people are willing to costly signal to others and to themselves that they possess prosocial traits, in order to maintain a positive social and self-image, respectively (see Bénabou and Tirole, 2006; Bénabou et al., 2018). Regarding social image concerns, it has been shown that an increase in one's observability increases generous and cooperative behavior, robustly confirming that people care about their prosocial reputation (Andreoni and Petrie, 2004; Ariely et al., 2009; Andreoni and Bernheim, 2009; Böhm and Regner, 2013; Bašić et al., 2018). Regarding

self-image concerns, the literature is more recent, nevertheless, the evidence is ample. Several studies use the willful ignorance paradigm, where subjects have the opportunity of increasing their payoff while potentially imposing a low payoff to another subject, but can choose to stay ignorant about the consequences to the other's payoff. It was shown that subjects become more selfish when they can stay ignorant, indicating that subjects avoid negative self-updating and try to maintain a positive self-image (Dana et al., 2007; Grossman, 2014; Grossman and van der Weele, 2017). More recently, Falk (2017) and Bašić et al. (2018) use an exogenous manipulation of self-image concerns where they increase the self-awareness of subjects and show that people start acting more morally and more generously, respectively. We add to this literature, as we shed light on the roots of these two fundamental drivers of prosocial behavior. So far, only a few studies investigate related questions with children. Engelmann et al. (2012) show that 5 year-old children steal less when being observed by another child, but find only weak evidence that observation influences their willingness to help. In another study, costly sharing was found to be influenced by an observer at the age of 5, but only when the observer was also able to give some resources to the decision maker immediately after the sharing decision (Engelmann et al., 2013). Although it was speculated that caring about one's reputation should become relevant at the age of 7-10, the evidence for this is still missing (Warneken, 2018).[1] We contribute to this important question with our experiment. We find only weak evidence that image concerns matter when we look at our entire subject pool, but find strong evidence that they affect boys and not girls, and especially strong evidence when we look at the 7-10 year-old boys. Their age is in line with the prediction of Warneken (2018). Moreover, given the strong gender difference, the finding gives important insights to the interdisciplinary literature on asymmetric gender effects of costly signaling (Trivers, 1972; Amotz Zahavi and Avishag Zahavi, 1999; E. A. Smith and Bird, 2000; Böhm and Regner, 2013; Van Vugt and Iredale, 2013; Raihani and S. Smith, 2015).

The most prominent explanation why men should be more inclined to signal prosocial traits relies on sexual selection theory which implies that women bear higher minimal parental investments (see Trivers, 1972), and hence are more selective when finding a mate. Given that a costly prosocial action can signal i) one's quality in possessing wealth, and ii) the willingness to share that wealth, or simply, one's propensity to care about others (see Amotz Zahavi and Avishag

[1] Warneken (2018) argues that behavior driven by concerns for reputation involves several complex processes. It requires higher-level social-cognitive reasoning and, likely, much experience, implying it should not play an effect in early childhood. Importantly, children need to understand what type of opinion others form of them, which coincides with understanding impression management, a process where people try to influence the perceptions of others. This reasoning seems to develop roughly in the period from 7 to 10 years of age (Aloise-Young, 1993; Banerjee, 2002; Hill and Pillow, 2006).

Zahavi, 1999; E. A. Smith and Bird, 2000), evolution arguably favored men to be more likely to signal these qualities as a mating strategy. This proposition found support in adults, as it was shown that in public settings, men increase their costly prosocial behavior more than women (Böhm and Regner, 2013),[2] and the increase is driven by men trying to impress attractive women (Van Vugt and Iredale, 2013; Raihani and S. Smith, 2015). Our findings contribute to this literature by shedding light on the roots of this gender asymmetry. In line with the predictions, we find that boys react to increased observability more than girls (who do not react), and offer evidence that the roots of this gender difference lie early in human development. We also offer suggestive evidence that this difference cannot be explained by a gender difference in understanding of normative behavior, but it appears to be a true difference in preference for appearing prosocial. Interestingly, a similar gender pattern can be observed in our *Self-image* treatment, where we find that only boys become more generous. This is in line with the findings of Bašić et al. (2018), who show that when exposed to increased self-awareness, only adult men exhibit stronger generosity in a dictator game, while women do not. We discuss the potential mechanism behind the self-image gender asymmetry in the Discussion and conclusion section (4.4). Finally, our study also connects to the more general literature on development of prosocial behavior. One of the most common findings in the literature is that children become more prosocial as they get older (Harbaugh et al., 2003; Benenson et al., 2007; Fehr et al., 2008, 2013; Bauer et al., 2014; Kosse et al., forthcoming). We replicate this finding in our *Control* treatment, as we observe that children become more generous with age.

The remainder of the paper is structured as follows. In Section 4.2, we describe the experimental design and procedure. In Section 4.3, we report the results of our study. In Section 4.4, we discuss and conclude the paper.

4.2 Experimental design and procedure

We divide the following section into two subsections. In the first subsection, we explain the experimental design. In the second subsection, we describe the experimental procedure and the subject pool.

4.2.1 Experimental design

To study prosocial behavior we focus on generosity, a fundamental facet of human prosociality. For this purpose, subjects played a dictator game, first in a role of a dictator, and then in a role of a recipient, where they were matched to a different person. They were informed about the second game only after the first

[2] Bašić et al. (2018) find a similar but insignificant pattern.

one.[3] As dictators, subjects had to divide 10 tokens between themselves and their paired recipient who was another student of their own age. Each token in the dictator game was worth a certain amount of Croatian kunas.

Subjects made their decisions on computers. Given their young age, to ensure comprehension, the verbal instructions were given in one-to-one sessions with an experimenter (for experimental instructions see Appendix 4.B). Following the dictator game, subjects participated in a short cognitive ability test. To obtain a quick and precise measure of cognitive ability, we implemented the symbol correspondence test (Dohmen et al., 2010). Subjects were confronted with 9 unfamiliar symbols, each paired to a different digit, from 1 to 9. The pairing was presented in a table at the top of the screen. Their task was to input the correct digit as the symbol would appear on the screen. After each input a new symbol would appear. They had 90 seconds to input as many correct digits as possible. After the test, the subjects were asked about basic socio-demographics, pocket allowance, and whether they had heard about the content of the experiment before. Moreover, they were given experimentally-validated survey questions for eliciting time preferences (Falk et al. (2016), see Appendix 4.A for a detailed description). Before they left the experimental session they were explicitly instructed not to talk about the study with students who had not yet participated.

We designed three different treatments. In the *Self-image* treatment, we exogenously manipulated self-awareness, raising attention to the congruency of inner standards of behavior and one's current behavior, and consequently, increasing the salience of self-image concerns.[4] In order to increase self-awareness, we exposed the subjects to their own "self-image" as in Falk (2017). In particular, from the moment subjects sat in front of the computer, they were facing a real-time video of their face, which was captured by a camera on the top of their computer screen. The camera was positioned in a way that subjects could not evade its visual field. The video was prominently placed in the middle-upper part of their computer screen, while the decision options were placed bellow it (see Figure 4.2 in Appendix). As the subjects sat at their computer, their experi-

[3] In order to ensure a clean comparison of behavior across age, we informed the dictators that they participated in a second dictator game as recipients after the first game, during the payoff procedure. Otherwise, children might form expectations about what they will receive from dictators, and the capabilities for such a cognitively complex reasoning might vary across age, giving rise to a potential confound. Even more, if children have correct expectations about what they could receive, the expectations would increase with age as older dictators are more generous. Together with the information about being a recipient, during the payoff, subjects were again informed about the rules of the game and were told their paired dictator's decision. Subjects were informed that they were not matched to the same person as when they played in the role of a dictator.

[4] For the concept of self-awareness see objective self-awareness theory (Duval and Wicklund, 1972), and empirical studies supporting it (Duval and Wicklund, 1973; Carver, 1975; Diener and Wallbom, 1976; Duval et al., 1979; Beaman et al., 1979).

menters explained that the video was not being recorded. The video was turned off immediately after the dictator game.

In the *Social image* treatment, we exogenously manipulated observability by exposing participants to the observation of another subject, i.e., their observer, as in Bašić et al. (2018). The dictators in *Social image* treatment were facing the same procedures and decision as in *Self-image*, but instead of seeing their face in the video, they saw their observer's face. The observers were dictators' classmates who were seated in front of another computer. Observers saw i) the video footage of their paired dictator's face in real time, and ii) his or her decision screen in real time. Hence, both the observer and the dictator saw each other, and additionally, the observer saw the dictator's decision screen. The observers had only one task during the dictator game, to observe the actions of their paired dictator. The observers also received instructions about the rules of the dictator game by their own experimenter. All these details were common knowledge. Both dictators and observers were informed that the video is not being recorded. Moreover, they were instructed that any type of communication through the video was strictly forbidden. Together with their task to observe the dictator, observers also answered the following questions: i) "what do you think is fair - how many tokens should the student whom you see on the screen give to the other student", and ii) "if the student whom you see on the screen gave 2 tokens to the other student, how fair would it be". They answered the latter question using a 5 point likert scale ranging from "very unfair" to "very fair". We use the two questions as proxies for i) children's understanding of what an appropriate and normative behavior is, and ii) their judgment about a selfish deviation from the normative behavior.[5] Furthermore, together with the cognitive ability task and the questionnaire at the end of the session, both observers and dictators had to indicate if any of them tried to communicate through the video, and if so, in which way. The observers earned an equivalent of 5 tokens in Croatian kunas for their role of observer. To make average payoff identical across different roles in the game, observers additionally earned an equivalent of 5 tokens in Croatian kunas for the additional questions they had to answer. Analogously to the dictators, they were informed about the additional payment during the payoff procedure.

To design a comparable control treatment, we addressed two concerns. First, in both our treatments, subjects were exposed to a video on their screen. Focusing on a video could drain cognitive resources, which in turn might reduce the

[5] The key concern in this elicitation was to ensure that even 7 year-old children can fully comprehend and answer our questions. For that reason, we do not differentiate between personal or social norms, and we do not ask for the children's opinions of what their peers might think the norm is (see Krupka and Weber, 2013). Moreover, we choose the word "fair" as it is akin to the words "appropriate" and "moral", which are commonly used in order to elicit injunctive social norms (see Krupka and Weber, 2013), but is more understandable for children.

ability of subjects to restrain their selfish impulses and would bias our results (Gino et al., 2011; Achtziger et al., 2015).[6] Second, in both treatments subjects saw a person on their screen who was looking at them: themselves in the *Self-image* treatment, and observer in the *Social image* treatment. This bears potential of influencing one's behavior, as studies indicate that a simple social cue, such as a pair of observing eyes can provoke a feeling of observation and influence subjects to act more prosocially (Bateson et al., 2006; Rigdon et al., 2009).[7] To address these two concerns, our *Control* treatment is identical to the other two treatments, except that instead of seeing themselves or their observer, subjects saw a neutral mute video of another child looking at them. For the video, we chose a short clip of a TV show in which children present news. This ensured that the child in the video is constantly looking at the camera and is acting in a neutral manner, not transmitting any tendentious associations.[8] Moreover, this allowed us to choose two similar videos to better match the age of the child in the video to the age of the subjects: one video was used with subjects in 2nd and 4th grade, and another with subjects in 6th and 8th grade (see Figures 4.3 and 4.4 in Appendix).[9] It was clear that the video is an old prerecorded video, which was also explicitly mentioned by the experimenter.

4.2.2 Procedure and subject pool

The study was conducted in Samobor, a town in central Croatia with an approximate population of 38,000. The participants were students in 2nd, 4th, 6th and 8th grade of elementary school, and were 7-8, 9-10, 11-12 and 13-14 years old, respectively.[10] For the main analysis we group children into two age cohorts:

[6] Note, however, that several studies suggest that the fundamental human impulse is to react prosocially; hence, in this respect, existence of a video could favor prosocial behavior (Rand et al., 2012; Schulz et al., 2014).

[7] Note, however, that some studies do not find an effect of social cues (Fehr and Schneider, 2010; Lamba and Mace, 2010), and question the validity of previous findings (Carbon and Hesslinger, 2011).

[8] Notice that our choice of the control treatment makes any potential effect of cognitive depletion or social cues constant across all three treatments. Hence, any behavioral difference caused by our two treatments can be attributed to self-awareness or observability, respectively.

[9] If social cues matter, videos with persons of different age could potentially cause different effects, e.g., a person older than the subject might trigger stronger willingness to act in line of societal standards. Hence, we took special care in choosing two appropriate and age-fitting videos. Note, however, that the change of the video has no effect on children's behavior. Specifically, in a regression model, video has no effect on dictator's giving when taking age into account, and the age coefficient stays very similar when excluding the video variable (see Table 4.4 in Appendix).

[10] Our sample contains few exceptions where subjects are slightly older or younger than their classmates. Also, one subject was 15 years of age; hence, we included the subject in the 11-14 year-old age cohort.

younger children (7-10y) and older children (11-14y), but we further analyze the data on a more detailed level using exact age when appropriate.[11]

We recruited our subjects from 2 elementary schools.[12] Specifically, we invited all students from all classes in 2nd, 4th, 6th and 8th grade to participate. Altogether, there were 30 participating classes across the two schools. 73% of invited students participated in the experiment, which resulted in 494 subjects (52% girls). 129 students participated in the *Control* treatment, 117 in the *Self-image* treatment, and 248 in the *Social image* treatment: 124 in the role of dictators and 124 in the role of recipients, respectively. Sessions were organized during regular school hours. Both schools ensured a private room for us in which we set up our lab in the field. Our lab was equipped with four laptops, each placed in a separate cubicle (see Figure 4.5 in Appendix). We conducted within-session randomization. In particular, there was one computer designated for *Self-image* treatment, one for *Control* treatment, and two for *Social image* treatment: one for observer and one for dictator. During the experiment, we balanced the experimenter assignments across different treatments in each school at the grade level to minimize any experimenter fixed effects.

At the beginning of each session an experimenter would arrive to a class which was participating in the study, and would randomly choose four subjects and escort them to the experimental room. One of the four experimenters would then assign them to different treatments while trying to ensure that children of different gender go to different treatments in comparison to the former session, ensuring balanced gender assignment across treatments. Before starting with the verbal instructions, each subject received a unique code. It was explained to the children that their choices are fully anonymous, and that instead of their name, only their code will be used during the study. Following that, the experimenters would proceed with the verbal experimental instructions for the respective treatments. During the explanation of the dictator game, subjects were informed that the recipient was a student of their own age but was not from their class. They were also informed that the recipients will know the rules

[11] The age split between younger and older children represents the average age of the children's age span, but more importantly, i) 11-12 years of age represents an important shift in children development, as it is the usual age when children enter puberty, and ii) recent work proposes that 7-10 is the age when reputation should start being an important factor (Warneken, 2018).

[12] The choice and the location of the two schools minimizes potential selection issues. Specifically, the schools participating in the study represent two out of three elementary schools in the town; there exists no clear difference in the quality of the three schools, and the children primarily enroll in a specific elementary school based on the proximity of their home. Furthermore, the probability of a child going to a different town to elementary school is negligible, similarly as the probability of moving to a different town due to perceived differences in school quality. The latter is extremely uncommon in Croatian society. Finally, note that elementary school is obligatory in Croatia, and the self-selection into different tracks occurs when students finish elementary school and choose between different high schools.

of the dictator game, and will know their final decision in the game. At the end of the instructions, subjects were asked to answer the control questions. In case a subject would not succeed in answering the control questions, the experimenter would explain the main aspects of the game again and repeat the control questions. Those who did not manage to answer all of the control questions continued their participation; however, we excluded them from the subsequent analysis. Out of 371 dictators, only 5 did not manage to answer the control questions.[13] To minimize any influence on the subject's behavior, the experimenter moved away from the computer during the dictator game, and sat on a chair from where neither the subject nor the subject's screen could be observed. Similarly, the experimenter turned away during the cognitive ability test to avoid any unnecessary pressure on a subject's performance.

To keep marginal incentives comparable, we adjusted the token values according to the grade. To estimate these values, we used a previously collected dataset of weekly pocket allowance from 3 Croatian schools, and adjusted the token values accordingly (see Table 4.5 in Appendix). 2nd grade children received 1 kuna for a token, 4th grade children received 1.5 kunas for a token, 6th grade children received 2.3 kunas for a token, and 8th grade children received 3.5 kunas for a token. 1 kuna = 0.145 US dollars at the time of the experiment. Alongside the payments from the game, children earned an additional 2 tokens worth of kunas for their participation in the study. The payment was given in a sealed envelope within two weeks after the experiment. The experiment was approved by the Croatian Ministry of Science and Education, the principals of both schools, and by the by Ethics Committee at the University of Bonn. Each parent gave written consent for their child to participate after learning about the nature of the study and the possible consequences. Each child gave verbal assent at the beginning of the experimental session to the experimenter.

4.3 Results

We divide our results section into two subsections. In the first subsection, we analyze the behavior of the entire sample. In the second subsection, we analyze and compare the behavior across genders.

4.3.1 Behavior of the entire sample

We first check if the treatments had an effect on the entire sample. We report OLS regression models in Table 4.1 where we regress the amount given by dictators on dummy variables for *Self-image* and *Social image* treatment and addi-

[13] Our findings stay robust when we include these 5 subjects in the analysis.

Table 4.1. OLS regressions estimates of treatment effects

Variables	Dependent variable: given amount of tokens					
	7-14y		7-10y		11-14y	
	(1)	(2)	(3)	(4)	(5)	(6)
Self-image	0.246	0.255	0.613*	0.458	-0.070	0.014
	(0.234)	(0.227)	(0.338)	(0.344)	(0.320)	(0.313)
Social image	0.267	0.422*	0.495	0.661*	0.058	0.291
	(0.244)	(0.245)	(0.373)	(0.367)	(0.318)	(0.331)
Control variables	No	Yes	No	Yes	No	Yes
Constant	3.180***	0.221	2.746***	1.364	3.551***	-0.131
	(0.165)	(2.156)	(0.237)	(3.545)	(0.223)	(2.507)
Observations	366	366	166	166	200	200
R-squared	0.004	0.078	0.020	0.095	0.001	0.121

Notes: The table presents OLS regressions using given amount of tokens as the dependent variable. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Control variables include daily pocket allowance, standardized measure of patience, number of siblings, child's age, cognitive abilities normalized at the age cohort level, and dummy variables for gender (1 if girl), hearing any information about the experiment before the session, attempt of communication from dictator towards observer, attempt of communication from observer towards dictator, and school fixed effects.

tional control variables.[14] Model (1) reveals that neither *Self* nor *Social image* treatment have a significant effect on the given amount. Model (2) reveals that if we include control variables, the *Self-image* treatment remains insignificant, while the *Social image* treatment shows weak indications of a positive effect, as the coefficient becomes significant at a 10% level.

Next we check if there is a development pattern of behavior across age, and divide our sample into two age groups: younger children (7-10y) and older children (11-14y). First, we observe that 11-14 year-old children give on average 3.551 tokens in the *Control* treatment (constant in Table 4.1, Column (5)), which is significantly more than the 2.746 tokens that 7-10 year-old children give (constant in Table 4.1, Column (3); two-sided t-test, $p = 0.015$, N(7-10y)=59, N(11-14y)=69). The same relation is observed if we regress giving on age of child, and also, if we include control variables (see Table 4.7 in Ap-

[14] The findings stay robust when using tobit estimations (see Table 4.6 in Appendix).

pendix).[15] On average, an increase of age by one year increases giving by 0.212 tokens. Next, we look at the linear regression models in Table 4.1 where we regress giving on the two treatments, separately for each age cohort. For 7-10 year-old children, we find a weakly significant effect of *Self-image* treatment; however, the significance disappears when including control variables (Models (3) and (4)). The opposite holds true for *Social image* treatment, as we observe significance with control variables but not without. Turning to 11-14 year-old children, we observe no significance for any of the two treatments (Models (5) and (6)).

Altogether, while we find slight indications that *Self* and *Social image* treatment cause an effect on behavior, the effects are small and not robust. This holds true for the entire sample, and also when we check for different age cohorts.

4.3.2 Behavior across genders

Next, to investigate potential gender differences proposed by evolutionary theories, we divide the sample across genders, and repeat the analysis separately for boys and girls in Table 4.2.[16] Models (1) and (2) show the effect of *Self-image* and *Social image* treatment on the entire sample of boys. We observe a significant effect of the *Social image* treatment, as boys increase their giving by 0.747 tokens, which corresponds to an increase of 25.6%. Similarly, we observe a significant effect of the *Self-image* treatment (at 10% level), as boys increase their giving by 0.553 tokens, which corresponds to an increase of 18.9%. Both effects remain significant when including control variables. Next, we focus on the behavior of girls (Models (3) and (4)). In contrast to boys, girls do not significantly react to any of the two manipulation treatments. To further investigate this difference in behavior across the two genders, we compare if there is a baseline difference in behavior without any manipulation. Although boys give less on average, we find no significant difference (2.919 vs 3.424, constants in Table 4.2, Columns (1) and (3); two-sided t-test, $p = 0.128$, N(boys)= 62, N(girl)=66). Second, we check for the interaction effects between gender and treatment (Models (5) and (6)). We find that when being observed, boys' giving increases significantly more than girls' giving (at 10% level), and this finding stays robust when including control variables. Moreover, we also find a similar gender pattern for the *Self-image* treatment, but the interaction effect is not significant.

[15] Five children from 2nd grade did not know when they were born, and we did not manage to ex-post find out what their age was; hence, we use the mean age of children within their grade level for their age.

[16] The findings stay robust when using tobit estimations (see Table 4.8 in Appendix).

Table 4.2. OLS regressions estimates of treatments effects by gender

Variables	Dependent variable: given amount of tokens						
	Boys 7-14y		Girls 7-14y		Boys and girls 7-14y		Boys 7-10y
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Self-image	0.553*	0.565*	-0.041	-0.020	0.553*	0.551*	1.202***
	(0.323)	(0.317)	(0.337)	(0.335)	(0.323)	(0.312)	(0.439)
Social image	0.747**	0.877**	-0.208	-0.042	0.747**	0.844**	1.066**
	(0.334)	(0.337)	(0.353)	(0.354)	(0.334)	(0.326)	(0.470)
Girl (=1)					0.505	0.437	
					(0.328)	(0.315)	
*Self-image**Girl					-0.594	-0.574	
					(0.467)	(0.454)	
*Social image**Girl					-0.955*	-0.829*	
					(0.486)	(0.464)	
Control variables	No	Yes	No	Yes	No	Yes	No
Constant	2.919***	-1.591	3.424***	1.816	2.919***	-0.053	2.452***
	(0.222)	(2.528)	(0.242)	(3.158)	(0.222)	(2.172)	(0.293)
Observations	180	180	186	186	366	366	86
R-squared	0.030	0.126	0.002	0.075	0.016	0.087	0.090
	Boys 7-10y	Boys 11-14y		Girls 7-10y		Girls 11-14y	
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Self-image	1.093**	-0.077	-0.057	0.003	-0.082	-0.048	0.063
	(0.452)	(0.462)	(0.486)	(0.512)	(0.554)	(0.445)	(0.436)
Social image	1.055**	0.407	0.652	-0.151	0.127	-0.256	-0.006
	(0.456)	(0.469)	(0.496)	(0.588)	(0.582)	(0.434)	(0.451)
Control variables	Yes	No	Yes	No	Yes	No	Yes
Constant	2.049	3.387***	0.251	3.071***	0.363	3.684***	0.305
	(3.386)	(0.317)	(5.483)	(0.375)	(6.330)	(0.314)	(2.561)
Observations	86	94	94	80	80	106	106
R-squared	0.247	0.014	0.130	0.001	0.070	0.004	0.150

Notes: The table presents OLS regressions using given amount of tokens as the dependent variable. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Control variables include daily pocket allowance, standardized measure of patience, number of siblings, child's age, cognitive abilities normalized at the age cohort level, and dummy variables for hearing any information about the experiment before the session, attempt of communication from dictator towards observer, attempt of communication from observer towards dictator, and school fixed effects.

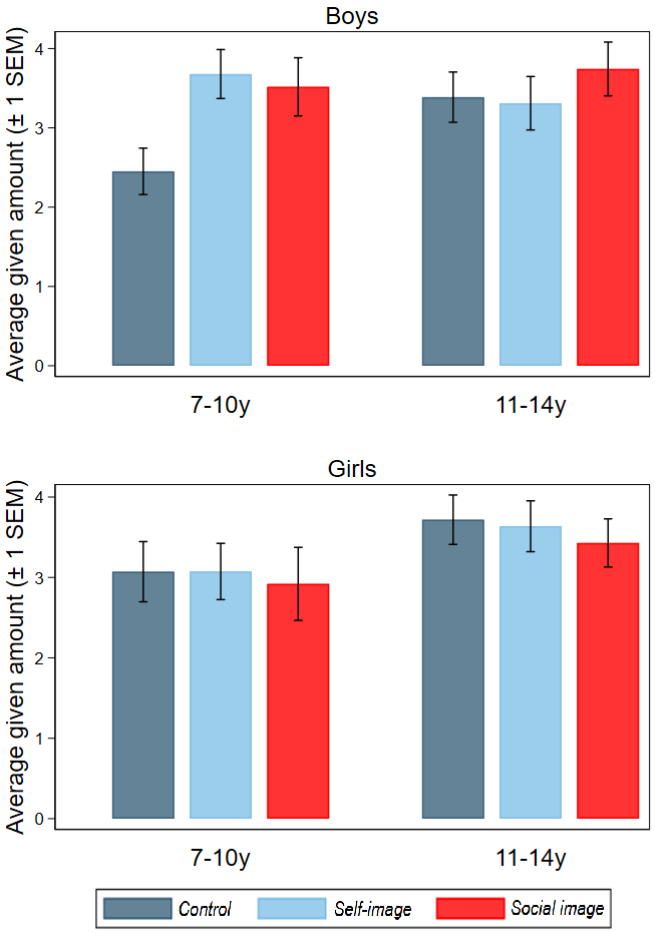

Figure 4.1. Average given amount of tokens for 7-10 and 11-14 year-old children separated by gender: boys (upper panel) and girls (lower panel)

Next, we investigate the development pattern of dictator behavior across age for the two genders. To visualize our findings, in Figure 4.1 we present the behavior for boys (upper panel) and girls (lower panel) for 7-10 year-old children and 11-14 year old-children. We first examine the behavior of 7-10 year-old boys (Table 4.2, Models (7) and (8)). We observe that young boys exhibit a significant reaction to both treatments, as they increase their giving by 1.202 tokens in the *Self-image* treatment (increase of 49.0%), and by 1.066 tokens in the *Social image* treatment (increase of 43.5%). The findings stay robust when controlling for additional variables. Older boys (11-14y), however, exhibit much weaker reactions to the two image treatments (Models (9) and (10)). Their giving in the *Self-image* treatment stays similar to the *Control* treatment, while

their giving in the *Social image* treatment increases by 0.407 tokens, but the increase is not significant. If we look at the behavior of girls (Models (11, 12, 13) and (14)), we see a stark contrast in comparison to the behavior of boys. Girls do not exhibit any reaction to the image treatments. None of the comparisons in 7-10 or 11-14 year-old girls shows significance. Finally, we also check if in the *Social image* treatment gender of the observer matters. Specifically, it could be that boys react more when girls are observing them (see Van Vugt and Iredale, 2013; Raihani and S. Smith, 2015). We do not find any evidence in support of this (see Table 4.9 in Appendix).

Altogether, we observe that in the overall sample, boys react positively to both *Self-image* treatment and *Social image* treatment, while girls do not react to any of the two treatments. Moreover, the reaction to the *Social image* treatment is significantly stronger in boys than in girls. Looking at the age development, we observe a significant effect of the two image treatments in younger boys (7-10y) but not in older boys (11-14y).

Next, to better understand the mechanisms behind the marked gender difference, we investigate if this can be explained by differences in understanding of normative behavior across genders. In particular, if boys think that the fairest transfer is higher than girls, one could argue that such normative valuation difference can drive the behavioral difference between the two genders, as the equivalent action might indicate selfishness in the view of boys, but not girls. Similarly, it could be that a certain violation of normative behavior is viewed as less appropriate by boys than by girls.[17] To investigate this conjecture, in Table 4.3 we present the average answers on questions about the fairest transfer in the game and the perceived fairness of giving 2 tokens, which were answered by the observers. The findings reveal that both questions about fairness do not differ across genders in any of the age cohorts or the entire sample, suggesting that the understanding of normative behavior cannot explain the behavioral gender difference we observe. Furthermore, the table reveals another important observation. For the 11-14 year-old children, the average opinion on the fairest choice comes fairly close to the children's actual behavior, suggesting a potential reason why we do not observe any effect of our two image treatments on behavior of the older children, at least in the case of boys who react to our treatments when they are younger (behavior=3.387 vs fairest choice=4.303 for 11-14 year-old boys, behavior=3.684 vs fairest choice=4.865 for 11-14 year-old girls).

[17] This connects to recent signaling models (e.g., Bénabou and Tirole, 2006), which imply the following: Falling short from existing social norms stigmatizes a person and negatively impacts his or her reputation, and also, an action which is below the social norm infers in expectation an exact (low) level of person's type, i.e., person's preference for generosity. Hence, if i) understanding of a normative behavior or ii) the judgment about the selfish deviation from a normative behavior would differ between the two genders, this might provoke asymmetric reactions to increased image concerns.

Table 4.3. Perceived fairest transfer and perceived fairness of giving 2 tokens

	7-14y		7-10y		11-14y	
	Fairest choice	Giving 2 tokens	Fairest choice	Giving 2 tokens	Fairest choice	Giving 2 tokens
Boys	4.649	2.596	5.125	2.459	4.303	2.700
	(0.215)	(0.143)	(0.423)	(0.225)	(0.192)	(0.187)
Girls	4.939	2.561	5.035	2.621	4.865	2.514
	(0.089)	(0.119)	(0.105)	(0.195)	(0.135)	(0.148)

Notes: ***p<0.01, **p<0.05, *p<0.1 for two-sided t-test comparisons between boys and girls for a given age cohort, *sd* in parentheses. The numbers represent the average answers of observers on following questions: "what do you think is fair - how many tokens should the student whom you see on the screen give to the other student", and "if the student whom you see on the screen gave 2 tokens to the other student, how fair would it be", separated by gender and age cohort.

4.4 Discussion and conclusion

In this study we show that both self and social image concerns are an important driver of prosocial behavior in children and adolescents (7-14y), however, only for boys. In our *Social image* treatment, we observe that boys react to increased observability, significantly stronger than girls, who do not react. This stark gender result supports evolutionary theories which suggest that men are more likely to signal prosocial traits (see Trivers, 1972; Amotz Zahavi and Avishag Zahavi, 1999; E. A. Smith and Bird, 2000; Raihani and S. Smith, 2015). Importantly, our results show that the roots of this gender asymmetry lie early in human development. We do not find that boys react more to girl observers as one might expect from the mentioned theories; however, studies that find such a result in adults use attractive women as observers and find that the decision maker's behavior is closely connected to the subjective perception of the attractiveness of the female observer (Van Vugt and Iredale, 2013; Raihani and S. Smith, 2015). This approach was unfeasible to implement with our young subject pool.

Similarly to the gender asymmetry we observe in the *Social image* treatment, we also observe that only boys react to our *Self-image* treatment. It is possible, as it could be explained by signaling models, that boys care more about their image than girls, regardless of the image domain. Connected to this, Von Hippel and Trivers (2011) suggest that when people need to convince others about a certain private quality, they engage in self-deception and convince themselves,

as it avoids the need for conscious deception. Schwardmann and van der Weele (2017) report causal evidence supporting this hypothesis, and show that such self-deception makes people more persuasive when convincing others. If so, the fact that boys are self-signaling prosocial traits could help them credibly convince others that they are prosocial when the chance arises, e.g., in our *Social image* treatment. Alternatively, boys' high self-image concerns do not need to be connected to their high social image concerns. Bašić et al. (2018) find that women are more prosocial than men, but when confronted with a self-awareness manipulation, only men increase their prosocial behavior, and they reach a similar level of prosocial behavior as women. The authors offer evidence that women are generally more self-aware than men, and hence might already be strongly focused on their inner standards of behavior without the self-awareness manipulation. While we do not find a significant difference in behavior across genders in our *Control* treatment, the behavioral patterns are in line with this potential explanation, suggesting that indeed, the fact that only boys react to the self-awareness manipulation could be due to girls' higher baseline self-awareness.

Our findings also offer insights for the question whether the observed gender difference is due to normative understanding or behavior, or is it a true difference in preference for image concerns. This is of special importance in our setup, as different genders might develop their normative understanding at a different pace, causing the gender findings we observe. Our findings suggest that the difference is not driven by differences in normative understanding, suggesting a true difference in preferences for image concerns. The normative valuation we elicit also connects to other important observations. Specifically, when dividing the sample in younger and older subjects, we observe the treatment effects only in younger boys. The cause of this is an increase of dictator's giving with age, which is offset by a stagnating level of giving in the two image treatments across younger and older boys. A reasonable explanation to this observation is the fact that in the *Control* treatment older boys act very close to what they think is fair. If in the image treatments subjects just want to appear fair, they again will act in a similar manner. Importantly, this might suggest that image concerns can cause an effect as long as the inner prosocial preferences are sufficiently underdeveloped and leave scope for an increase, as we observe in younger boys. Alternative to this explanation, it is possible that at the age of 11-14, boys' concerns for image diminish. This, however, goes strongly against the common finding that in early adolescence children become more aware of, and increasingly concerned with, the opinion of others (see Vartanian, 2000). They start shaping their social behavior according to their peers (Steinberg and Silverberg, 1986), and are distressed if they feel excluded (Sebastian et al., 2010). Hence, while our results cannot dismiss this explanation, it seems implausible.

References

Achtziger, Anja, Carlos Alós-Ferrer, and Alexander K. Wagner (2015): "Money, depletion, and prosociality in the dictator game." *Journal of Neuroscience, Psychology, and Economics*, 8 (1), 1–14. [63]

Aloise-Young, Patricia A. (1993): "The development of self-presentation: Self-promotion in 6-to 10-year-old children." *Social Cognition*, 11 (2), 201–222. [59]

Andreoni, James and B. Douglas Bernheim (2009): "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects." *Econometrica*, 77 (5), 1607–1636. [58]

Andreoni, James and Ragan Petrie (2004): "Public goods experiments without confidentiality: A glimpse into fund-raising." *Journal of Public Economics*, 88 (7-8), 1605–1623. [57, 58]

Ariely, Dan, Anat Bracha, and Stephan Meier (2009): "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *The American Economic Review*, 99 (1), 544–555. [57, 58]

Banerjee, Robin (2002): "Audience effects on self-presentation in childhood." *Social Development*, 11 (4), 487–507. [59]

Bašić, Zvonimir, Armin Falk, and Simone Quercia (2018): "Self-image, social image and prosocial behavior." [58–60, 62, 72]

Bateson, Melissa, Daniel Nettle, and Gilbert Roberts (2006): "Cues of being watched enhance cooperation in a real-world setting." *Biology Letters*, 2 (3), 412–414. [63]

Bauer, Michal, Julie Chytilová, and Barbara Pertold-Gebicka (2014): "Parental background and other-regarding preferences in children." *Experimental Economics*, 17 (1), 24–46. [60]

Beaman, Arthur L., Bonnel Klentz, Edward Diener, and Soren Svanum (1979): "Self-awareness and transgression in children: Two field studies." *Journal of Personality and Social Psychology*, 37 (10), 1835–1846. [61]

Bénabou, Roland, Armin Falk, and Jean Tirole (2018): "Narratives, imperatives and moral reasoning." *Working Paper*. [58]

Bénabou, Roland and Jean Tirole (2006): "Incentives and prosocial behavior." *The American Economic Review*, 96 (5), 1652–1678. [58, 70]

Benenson, Joyce F., Joanna Pascoe, and Nicola Radmore (2007): "Children's altruistic behavior in the dictator game." *Evolution and Human Behavior*, 28 (3), 168–175. [60]

Böhm, Robert and Tobias Regner (2013): "Charitable giving among females and males: An empirical test of the competitive altruism hypothesis." *Journal of Bioeconomics*, 15 (3), 251–267. [58–60]

Carbon, Claus-Christian and Vera M. Hesslinger (2011): "Bateson et al.'s (2006) cues-of-being-watched paradigm revisited." *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*, 70 (4), 203–210. [63]

Carpenter, Jeffrey and Caitlin Knowles Myers (2010): "Why volunteer? Evidence on the role of altruism, image, and incentives." *Journal of Public Economics*, 94 (11-12), 911–920. [57]

Carver, Charles S. (1975): "Physical aggression as a function of objective self-awareness and attitudes toward punishment." *Journal of Experimental Social Psychology*, 11 (6), 510–519. [61]

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang (2007): "Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness." *Economic Theory*, 33 (1), 67–80. [59]

DellaVigna, Stefano, John A. List, and Ulrike Malmendier (2012): "Testing for altruism and social pressure in charitable giving." *The Quarterly Journal of Economics*, 127 (1), 1–56. [57]

Diener, Edward and Mark Wallbom (1976): "Effects of self-awareness on antinormative behavior." *Journal of Research in Personality*, 10 (1), 107–111. [61]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2010): "Are risk aversion and impatience related to cognitive ability?" *American Economic Review*, 100 (3), 1238–60. [61]

Duval, Shelley, Virginia H. Duval, and Robert Neely (1979): "Self-focus, felt responsibility, and helping behavior." *Journal of Personality and Social Psychology*, 37 (10), 1769–1779. [61]

Duval, Shelley and Robert A. Wicklund (1972): *A theory of objective self awareness*. Oxford: Academic Press. [61]

Duval, Shelley and Robert A. Wicklund (1973): "Effects of objective self-awareness on attribution of causality." *Journal of Experimental Social Psychology*, 9 (1), 17–31. [61]

Engelmann, Jan M., Esther Herrmann, and Michael Tomasello (2012): "Five-year olds, but not chimpanzees, attempt to manage their reputations." *PLoS One*, 7 (10), e48433. [59]

Engelmann, Jan M., Harriet Over, Esther Herrmann, and Michael Tomasello (2013): "Young children care more about their reputation with ingroup members and potential reciprocators." *Developmental Science*, 16 (6), 952–958. [59]

Falk, Armin (2017): "Facing yourself – A note on self-image." *CESifo Working Paper Series* (6428). [57, 59, 61]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde (2016): "The preference survey module: A validated instrument for measuring risk, time, and social preferences." *IZA Discussion Papers* (9674). [61, 76]

Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach (2008): "Egalitarianism in young children." *Nature*, 454, 1079–1083. [60]

Fehr, Ernst, Daniela Glätzle-Rützler, and Matthias Sutter (2013): "The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence." *European Economic Review*, 64, 369–383. [60]

Fehr, Ernst and Frédéric Schneider (2010): "Eyes are on us, but nobody cares: Are eye cues relevant for strong reciprocity?" *Proceedings of the Royal Society of London B: Biological Sciences*, 277 (1686), 1315–1323. [63]

Gino, Francesca, Maurice E. Schweitzer, Nicole L. Mead, and Dan Ariely (2011): "Unable to resist temptation: How self-control depletion promotes unethical behavior." *Organizational Behavior and Human Decision Processes*, 115 (2), 191–203. [63]

Grossman, Zachary (2014): "Strategic ignorance and the robustness of social preferences." *Management Science*, 60 (11), 2659–2665. [59]

Grossman, Zachary and Joël van der Weele (2017): "Self-image and willful ignorance in social decisions." *Journal of the European Economic Association*, 15 (1), 173–217. [57, 59]

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2009): "Cultural biases in economic exchange?" *The Quarterly Journal of Economics*, 124 (3), 1095–1131. [57]

Harbaugh, William T., Kate Krause, and Steven G. Liday Jr. (2003): "Bargaining by children." *University of Oregon Economics Working Paper* (2002-4). [60]

Hill, Valerie and Bradford H. Pillow (2006): "Children's understanding of reputations." *The Journal of Genetic Psychology*, 167 (2), 137–157. [59]

Kosse, Fabian, Thomas Deckers, Pia R. Pinger, Hannah Schildberg-Hörisch, and Falk Armin (forthcoming): "The formation of prosociality: Causal evidence on the role of social environment." *Journal of Political Economy*. [60]

Krupka, Erin L. and Roberto A. Weber (2013): "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association*, 11 (3), 495–524. [62]

Lamba, Shakti and Ruth Mace (2010): "People recognise when they are really anonymous in an economic game." *Evolution and Human Behavior*, 31 (4), 271–278. [63]

Raihani, Nichola J. and Sarah Smith (2015): "Competitive helping in online giving." *Current Biology*, 25 (9), 1183–1186. [58–60, 70, 71]

Rand, David G., Joshua D. Greene, and Martin A. Nowak (2012): "Spontaneous giving and calculated greed." *Nature*, 489, 427–430. [63]

Rigdon, Mary, Keiko Ishii, Motoki Watabe, and Shinobu Kitayama (2009): "Minimal social cues in the dictator game." *Journal of Economic Psychology*, 30 (3), 358–367. [63]

Schulz, Jonathan F., Urs Fischbacher, Christian Thöni, and Verena Utikal (2014): "Affect and fairness: Dictator games under cognitive load." *Journal of Economic Psychology*, 41, 77–87. [63]

Schwardmann, Peter and Joël van der Weele (2017): "Deception and self-deception." *Discussion Paper, Collaborative Research Center Transregio 190 - Rationality and Competition, München* (25). [72]

Sebastian, Catherine, Essi Viding, Kipling D. Williams, and Sarah-Jayne Blakemore (2010): "Social brain development and the affective consequences of ostracism in adolescence." *Brain and Cognition*, 72 (1), 134–145. [72]

Smith, Eric Alden and Rebecca L. Bliege Bird (2000): "Turtle hunting and tombstone opening: Public generosity as costly signaling." *Evolution and Human Behavior*, 21 (4), 245–261. [58–60, 71]

Steinberg, Laurence and Susan B. Silverberg (1986): "The vicissitudes of autonomy in early adolescence." *Child Development*, 57 (4), 841–851. [72]

Trivers, Robert (1972): "Parental investment and sexual selection." In *Sexual selection and the descent of man, 1871–1971*. Ed. by Bernard Campbell. Aldine Publishing Company, Chicago, 136–207. [58, 59, 71]

Van Vugt, Mark and Wendy Iredale (2013): "Men behaving nicely: Public goods as peacock tails." *British Journal of Psychology*, 104 (1), 3–13. [59, 60, 70, 71]

Vartanian, Lesa Rae (2000): "Revisiting the imaginary audience and personal fable constructs of adolescent egocentrism: A conceptual review." *Adolescence*, 35 (140), 639–662. [72]

Von Hippel, William and Robert Trivers (2011): "The evolution and psychology of self-deception." *Behavioral and Brain Sciences*, 34 (1), 1–56. [71]

Warneken, Felix (2018): "How children solve the two challenges of cooperation." *Annual Review of Psychology*, 69, 205–229. [58, 59, 64]

Zahavi, Amotz and Avishag Zahavi (1999): *The handicap principle: A missing piece of Darwin's puzzle*. Oxford University Press. [58, 59, 71]

Zak, Paul J. and Stephen Knack (2001): "Trust and growth." *The Economic Journal*, 111 (470), 295–321. [57]

Appendix 4.A Time preferences

For eliciting time preferences we used experimentally validated questions of Falk et al. (2016). We used one main question from the refined version of their preference module: "How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future". Given the complexity of the second suggested question in their preference module, we decided for an alternative question from their study as our second question of time preferences: "Do you often postpone things even though it would be better to take care of them right away". Both questions were answered on a 5 point likert scale. To ensure comprehension, experimenters were also providing examples, fitting to the respective question. To produce the measure of time preferences that we use in the study (patience), we added the answer on the first question to the reverse answer on the second question, and have standardized the measure.

Appendix 4.B Experimental instructions

The following section contains experimental instructions translated from Croatian.

4.B.1 Dictator: *Control* treatment

Hello, take a seat.

Please give me your parental consent. (*The experimenter turns the parental consent on the blank side and puts it on the table*.)

Ok, we can start. This is a study in which you will earn some money. The study consists of one game and some tasks. You will get (*value of two tokens adjusted to the grade level*) kunas just for participation, and additionally you can win some money in the game. Before we start with the study, I want to tell you that your participation is fully voluntary, and you can decide to quit at any time,

without any consequences. So first, I want to ask you: Do you want to start with the study? (*The experimenter writes down the subject's answer on the blank side of parental consent, and signs next to it.*)

You will get all the money that you earn in the study in a closed envelope within 2 weeks of today. During the study, we will use a code instead of your name. Your name will never be used. Only your code will be written on the envelope containing your money. (*The experimenter takes a deck of cards with codes written on them, and holds it so the cards are turned down. Following that, the experimenter takes the top card and shows it to the child.*) Here, this is your card with your code. I will type it now in the computer. (*Experimenter types the code and presses the button for next screen.*) Here, this is your code now. (*Experimenter returns the card to the child*.) Take good care of it, so you can get your money! Also write the code from the card in your notebook later, so in case you lose the card you will still know your code.

Ok, now you will play a game. Before you start, I will explain you the rules of the game. You will play the game on this computer. During the game, you will constantly see this old TV video playing on your screen.

Now I will explain you the rules. The game is being played in pairs of two. You are in a pair with one other player of similar age as you but not from your class. I will not tell you who he is neither will I tell the second player who you are. That will remain a secret.

The second player will get (*value of two tokens adjusted to the grade level*) kunas for participation, the same as you. How many kunas will the second player get in the game, that depends on you. The game is played with tokens. Each token is worth (*token value adjusted to the grade level*) kunas. That means that for every token that you are left with at the end of the game, you will get (*token value adjusted to the grade level*) kunas. Equivalently, the second player will also get (*token value adjusted to the grade level*) kunas for each token that he is left with at the end of the game.

At the beginning of the game you have 10 tokens. The second player has 0 tokens. Your task is to decide how many tokens you want to give to the second player. The second player cannot influence your decision, and at the end of the game he will only get the tokens that you give him. Also, the second player will know that you had 10 tokens at the beginning of the game, and he will know how many you left for yourself, and how many you gave him.

(*The experimenter takes a sheet with the printed decision screen.*) Here, on this picture you can see how your task will look on the computer screen. Your task is

to decide how many tokens you want to give to the second player. For example, if you want to give 0 tokens to the second player, and leave all 10 tokens for yourself, then you just click on this box, where it's written "0" - hence, you give him 0 tokens. If you want to give 5 tokens to the second player, and leave 5 tokens for yourself, then you click on this box, where it's written "5" - so, you give him 5 tokens. Or for example, if you want to give 2 tokens to the second player, and leave 8 tokens for yourself, then you click here where it's written "2". When you make your decision, you still have to click the "confirm" button to confirm your decision. When you'll be making your decision I will move away, so I won't find out how many tokens you gave to the second player. Do you have any questions?

Ok, now I will ask you several questions to see if you understood the game.

1. How much money is each token worth?
2. Who is the second player in the game?
3. How many tokens do you have, and how many tokens does the second player have at the beginning of the game?
4. What is your task in the game?

(*If the subject does not know how to answer one question, the experimenter answers it, and repeats the question at the end. If the subject does not know how to answer two questions, the experimenter shortly repeats the rules of the game and then again asks all the questions. If the subject again does not know how to answer one question, the experimenter answers it, and repeats it at the end. If the subject again fails to answer two questions, the experiment continues, and at the end of the study experimenter marks that the subject failed to answer the control questions.*)

Ok, I will move away now. When you are ready to play the game press this button in the lower-right corner and wait until the next screen appears, where you will decide how many tokens you want to give to the second player. Do you have any remaining questions? When you will be done, please raise your hand. When I see that you raised your hand, I will come back. (*The experimenter moves aside and sits on a chair. After the subject raises his hand, the experimenter comes back, turns off the video, and continues with the symbol correspondence test and the questionnaire. At the end of the study, the subject is told not to talk about the study with other students that did not yet participate*.)

4.B.2 Dictator: *Self-image* treatment

Hello, take a seat.

Please give me your parental consent. (*The experimenter turns the parental consent on the blank side and puts it on the table*.)

Ok, we can start. This is a study in which you will earn some money. The study consists of one game and some tasks. You will get (*value of two tokens adjusted to the grade level*) kunas just for participation, and additionally you can win some money in the game. Before we start with the study, I want to tell you that your participation is fully voluntary, and you can decide to quit at any time, without any consequences. So first, I want to ask you: Do you want to start with the study? (*The experimenter writes down the subject's answer on the blank side of parental consent, and signs next to it.*)

You will get all the money that you earn in the study in a closed envelope within 2 weeks of today. During the study, we will use a code instead of your name. Your name will never be used. Only your code will be written on the envelope containing your money. (*The experimenter takes a deck of cards with codes written on them, and holds it so the cards are turned down. Following that, the experimenter takes the top card and shows it to the child.*) Here, this is your card with your code. I will type it now in the computer. (*Experimenter types the code and presses the button for next screen.*) Here, this is your code now. (*Experimenter returns the card to the child*.) Take good care of it, so you can get your money! Also write the code from the card in your notebook later, so in case you lose the card you will still know your code.

Ok, now you will play a game. Before you start, I will explain you the rules of the game. You will play the game on this computer. During the game, you will constantly see your reflection in the screen. (*Experimenter waves his or her hand to show that it is a real-time video*). That is just your reflection through the camera - the video is not being recorded.

Now I will explain you the rules. The game is being played in pairs of two. You are in a pair with one other player of similar age as you but not from your class. I will not tell you who he is neither will I tell the second player who you are. That will remain a secret.

The second player will get (*value of two tokens adjusted to the grade level*) kunas for participation, the same as you. How many kunas will the second player get in the game, that depends on you. The game is played with tokens. Each token is worth (*token value adjusted to the grade level*) kunas. That means that for every token that you are left with at the end of the game, you will get (*token value adjusted to the grade level*) kunas. Equivalently, the second player will also get (*token value adjusted to the grade level*) kunas for each token that he is left with at the end of the game.

At the beginning of the game you have 10 tokens. The second player has 0 tokens. Your task is to decide how many tokens you want to give to the second player. The second player cannot influence your decision, and at the end of the game he will only get the tokens that you give him. Also, the second player will know that you had 10 tokens at the beginning of the game, and he will know how many you left for yourself, and how many you gave him.

(*The experimenter takes a sheet with the printed decision screen.*) Here, on this picture you can see how your task will look on the computer screen. Your task is to decide how many tokens you want to give to the second player. For example, if you want to give 0 tokens to the second player, and leave all 10 tokens for yourself, then you just click on this box, where it's written "0" - hence, you give him 0 tokens. If you want to give 5 tokens to the second player, and leave 5 tokens for yourself, then you click on this box, where it's written "5" - so, you give him 5 tokens. Or for example, if you want to give 2 tokens to the second player, and leave 8 tokens for yourself, then you click here where it's written "2". When you make your decision, you still have to click the "confirm" button to confirm your decision. When you'll be making your decision I will move away, so I won't find out how many tokens you gave to the second player. Do you have any questions?

Ok, now I will ask you several questions to see if you understood the game.

1. How much money is each token worth?
2. Who is the second player in the game?
3. How many tokens do you have, and how many tokens does the second player have at the beginning of the game?
4. What is your task in the game?

(*If the subject does not know how to answer one question, the experimenter answers it, and repeats the question at the end. If the subject does not know how to answer two questions, the experimenter shortly repeats the rules of the game and then again asks all the questions. If the subject again does not know how to answer one question, the experimenter answers it, and repeats it at the end. If the subject again fails to answer two questions, the experiment continues, and at the end of the study experimenter marks that the subject failed to answer the control questions.*)

Ok, I will move away now. When you are ready to play the game press this button in the lower-right corner and wait until the next screen appears, where you will decide how many tokens you want to give to the second player. Do you have any remaining questions? When you will be done, please raise your hand. When I see that you raised your hand, I will come back. (*The experimenter moves aside and sits on a chair. After the subject raises his or her hand, the experimenter*

comes back, turns off the video, and continues with the symbol correspondence test and the questionnaire. At the end of the study, the subject is told not to talk about the study with other students that did not yet participate.)

4.B.3 Dictator: *Social image* treatment

Hello, take a seat.

Please give me your parental consent. (*The experimenter turns the parental consent on the blank side and puts it on the table*.)

Ok, we can start. This is a study in which you will earn some money. The study consists of one game and some tasks. You will get (*value of two tokens adjusted to the grade level*) kunas just for participation, and additionally you can win some money in the game. Before we start with the study, I want to tell you that your participation is fully voluntary, and you can decide to quit at any time, without any consequences. So first, I want to ask you: Do you want to start with the study? (*The experimenter writes down the subject's answer on the blank side of parental consent, and signs next to it.*)

You will get all the money that you earn in the study in a closed envelope within 2 weeks of today. During the study, we will use a code instead of your name. Your name will never be used. Only your code will be written on the envelope containing your money. (*The experimenter takes a deck of cards with codes written on them, and holds it so the cards are turned down. Following that, the experimenter takes the top card and shows it to the child.*) Here, this is your card with your code. I will type it now in the computer. (*Experimenter types the code and presses the button for next screen.*) Here, this is your code now. (*Experimenter returns the card to the child*.) Take good care of it, so you can get your money! Also write the code from the card in your notebook later, so in case you lose the card you will still know your code.

Ok, now you will play a game. Before you start, I will explain you the rules of the game. You will play the game on this computer. As you see, you can observe another student from your class on the screen. Come on, wave to him / her so you can see that he / she also sees you. As you can see, you see the student in real-time. The video is not being recorded. The student that you see is your observer. Your observer sees you on his / her computer, but he / she also sees your screen. That means that he / she sees every move you make with your mouse, and is going to see your decisions during the game. His / her task is to observe the decisions you make during the game. Same as you, he / she will know the rules of the game. During the game, you cannot communicate with the observer, ok? That means that you cannot wave, make strange faces, etc.

Now I will explain you the rules. The game is being played in pairs of two. You are in a pair with one other player of similar age as you but not from your class. I will not tell you who he is neither will I tell the second player who you are. That will remain a secret.

Just to mention, the second player that is paired with you is not the observer that you see on the screen.

The second player will get (*value of two tokens adjusted to the grade level*) kunas for participation, the same as you. How many kunas will the second player get in the game, that depends on you. The game is played with tokens. Each token is worth (*token value adjusted to the grade level*) kunas. That means that for every token that you are left with at the end of the game, you will get (*token value adjusted to the grade level*) kunas. Equivalently, the second player will also get (*token value adjusted to the grade level*) kunas for each token that he is left with at the end of the game.

At the beginning of the game you have 10 tokens. The second player has 0 tokens. Your task is to decide how many tokens you want to give to the second player. The second player cannot influence your decision, and at the end of the game he will only get the tokens that you give him. Also, the second player will know that you had 10 tokens at the beginning of the game, and he will know how many you left for yourself, and how many you gave him.

(*The experimenter takes a sheet with the printed decision screen.*) Here, on this picture you can see how your task will look on the computer screen. Your task is to decide how many tokens you want to give to the second player. For example, if you want to give 0 tokens to the second player, and leave all 10 tokens for yourself, then you just click on this box, where it's written "0" - hence, you give him 0 tokens. If you want to give 5 tokens to the second player, and leave 5 tokens for yourself, then you click on this box, where it's written "5" - so, you give him 5 tokens. Or for example, if you want to give 2 tokens to the second player, and leave 8 tokens for yourself, then you click here where it's written "2". When you make your decision, you still have to click the "confirm" button to confirm your decision. When you'll be making your decision I will move away, so I won't find out how many tokens you gave to the second player. Do you have any questions?

Ok, now I will ask you several questions to see if you understood the game.

1. What is the observer's task in the game?
2. How much money is each token worth?
3. Who is the second player in the game?

4. How many tokens do you have, and how many tokens does the second player have at the beginning of the game?
5. What is your task in the game?

(*If the subject does not know how to answer one question, the experimenter answers it, and repeats the question at the end. If the subject does not know how to answer two questions, the experimenter shortly repeats the rules of the game and then again asks all the questions. If the subject again does not know how to answer one question, the experimenter answers it, and repeats it at the end. If the subject again fails to answer two questions, the experiment continues, and at the end of the study experimenter marks that the subject failed to answer the control questions.*)

Ok, I will move away now. When you are ready to play the game press this button in the lower-right corner and wait until the next screen appears, where you will decide how many tokens you want to give to the second player. Do you have any remaining questions? When you will be done, please raise your hand. When I see that you raised your hand, I will come back. (*The experimenter moves aside and sits on a chair. After the subject raises his or her hand, the experimenter comes back, turns off the video, and continues with the symbol correspondence test and the questionnaire. At the end of the study, the subject is told not to talk about the study with other students that did not yet participate*.)

4.B.4 Observer

Hello, take a seat.

Please give me your parental consent. (*The experimenter turns the parental consent on the blank side and puts it on the table*.)

Ok, we can start. This is a study in which you will earn some money. The study consists of one game and some tasks. You will get (*value of two tokens adjusted to the grade level*) kunas just for participation, and additionally you will earn some money for your role in the game. Before we start with the study, I want to tell you that your participation is fully voluntary, and you can decide to quit at any time, without any consequences. So first, I want to ask you, do you want start with the study? (*The experimenter writes down the subject's answer on the blank side of parental consent, and signs next to it.*)

You will get all the money that you earn in the study in a closed envelope within two weeks of today. During the study, we will use a code instead of your name. Your name will never be used. Only your code will be written on the envelope containing your money. (*The experimenter takes a deck of cards with codes written on them, and holds it so the cards are turned down. Following that,*

the experimenter takes the top card and shows it to the child.) Here, this is your card with your code. Take good care of it, so you can get your money! Also write the code from the card in your notebook later, so in case you lose the card you will still know your code.

Ok, now we can start with the study. You will participate in the study on this computer. You can see another student from your class on the screen. Let's wait until the student waves to you. (*Observer waits until the other student waves.*) Wave back to him / her, so he / she knows that you see him / her. As you can see, you see the student in real-time. The video is not being recorded. As you can see him / her, he / she can also see you. But in addition, you also see his / her screen. What you see here is actually the screen of the student that you are observing.

The student that you see will play one game. Your task is to observe what he / she will do in this game. You will see everything that he / she does here on the screen. You just have to wait and observe. You cannot interfere or communicate with the child that is playing the game, ok? That means that you cannot wave, make strange faces, etc. For your role of the observer you will get (*value of five tokens adjusted to the grade level*) kunas.

Now I will explain you the rules of the game that you will watch. The student that you are observing got 10 tokens. Each token is worth (*token value adjusted to the grade level*) kunas. That means that for each token that he / she is left with at the end of that game, he / she is going to get (*token value adjusted to the grade level*) kunas. The task of this student is to decide how many tokens he / she wants to give to one other student. He / she does not know who this student is, but he / she knows that this student is of similar age, but is from another class. The student you are observing can give as many tokens as he / she likes to the other student. That is his / her decision. The other student that will receive the tokens will not know who gave him the tokens. He will just know that there were 10 tokens to divide, and how many tokens he got.

(*The experimenter takes a sheet with the printed decision screen.*) Here, on this picture you can see how the task of this student will look on the computer screen. When he / she makes a decision how many tokens he / she wants to give to the other student, he / she will press the box with that number. For example, if he / she wants to give 3 tokens to the other student, and leave 7 tokens for himself / herself, he / she will choose the box with number "3" and press on it. Or, if he / she wants to give 1 token to the other student, and leave 9 tokens for himself / herself, he / she will choose the box with number "1". You should watch carefully which box the student will choose. Only you are going

to see that, since I will not be watching what he / she decides. Do you have any questions?

Ok, just to shortly repeat the rules, the student that you see on the screen got 10 tokens. Each token is worth (*token value adjusted to the grade level*) kunas. He / she must decide how many tokens he / she wants to give to one other, unknown student. Before you start with your role of observer I have two questions for you:

1. What do you think is fair - how many tokens should the student whom you see on the screen give to the other student?
2. If the student whom you see on the screen gave 2 tokens to the other student, how fair would it be? (*Experimenter writes down the answers*.)

Now you just have to wait a bit longer, until the student from the screen gets his / her instructions. When he / she is done with instructions, he / she will start with the game. I will leave now, but I will come back when the student you are observing finishes with the game. You just sit here and watch what he / she will do. (*The experimenter moves aside and sits on a chair. After the subject's paired dictator raises his or her hand, the experimenter comes back, turns off the video, types in the code that the subject received at the beginning of the study, and continues with the symbol correspondence test and the questionnaire. At the end of the study, the subject is told not to talk about the study with other students that did not yet participate*.)

Appendix 4.C Additional tables and figures

Table 4.4. OLS regressions estimates of the effect of video and age in *Control* treatment

Variables	Dependent variable: given amount of tokens	
	Control treatment	
	(1)	(2)
Age	0.212***	0.241*
	(0.071)	(0.145)
Older children video (=1)		-0.155
		(0.662)
Constant	0.900	0.664
	(0.778)	(1.276)
Observations	128	128
R-squared	0.067	0.067

Notes: The table presents OLS regressions using given amount of tokens as the dependent variable. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 4.5. Adjustment of the token value

Grade (age)	Average weekly pocket allowance	Increase	Token value	Increase
2nd (7-8y)	7.03 kn		1 kn	
4th (9-10y)	10.55 kn	+50%	1.5 kn	+43%
6th (11-12y)	15.55 kn	+48%	2.3 kn	+50%
8th (13-14y)	23.74 kn	+53%	3.5 kn	+53%

Notes: Average weekly pocket allowance was calculated based on a previously collected dataset from 3 Croatian schools. Token value represents the value of 1 token in our study (in Croatian kunas). Increase measures the relative increase in average weekly pocket allowance or token value for a grade in row x in comparison to row $x - 1$, respectively.

Table 4.6. Tobit regressions estimates of treatment effects

Variables	Dependent variable: given amount of tokens					
	7-14y		7-10y		11-14y	
	(1)	(2)	(3)	(4)	(5)	(6)
Self-image	0.322	0.330	0.732*	0.564	-0.025	0.054
	(0.256)	(0.244)	(0.374)	(0.367)	(0.344)	(0.328)
Social image	0.301	0.499*	0.529	0.773*	0.090	0.350
	(0.272)	(0.265)	(0.422)	(0.398)	(0.348)	(0.349)
Control variables	No	Yes	No	Yes	No	Yes
Constant	3.068***	-0.875	2.594***	-0.235	3.470***	-0.841
	(0.189)	(2.452)	(0.279)	(3.926)	(0.247)	(2.725)
Observations	366	366	166	166	200	200
Pseudo R-sq.	0.001	0.021	0.005	0.027	0.000	0.031

Notes: The table presents tobit regressions using given amount of tokens as the dependent variable with censoring at the amounts of 0 and 10. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Control variables include daily pocket allowance, standardized measure of patience, number of siblings, child's age, cognitive abilities normalized at the age cohort level, and dummy variables for gender (1 if girl), hearing any information about the experiment before the session, attempt of communication from dictator towards observer, attempt of communication from observer towards dictator, and school fixed effects.

Table 4.7. OLS regressions estimates of the effect of age in *Control* treatment

Variables	Dependent variable: given amount of tokens	
	Control treatment	
	(1)	(2)
Age	0.212***	0.157**
	(0.071)	(0.074)
Control variables	No	Yes
Constant	0.900	-0.430
	(0.778)	(1.666)
Observations	128	128
R-squared	0.067	0.157

Notes: The table presents OLS regressions using given amount of tokens as the dependent variable. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Control variables include daily pocket allowance, standardized measure of patience, number of siblings, cognitive abilities normalized at the age cohort level, and dummy variables for gender (1 if girl), hearing any information about the experiment before the session, and school fixed effects.

Table 4.8. Tobit regressions estimates of treatments effects by gender

Variables	Dependent variable: given amount of tokens						
	Boys 7-14y		Girls 7-14y		Boys and girls 7-14y		Boys 7-10y
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Self-image	0.652*	0.658**	0.009	0.030	0.658*	0.654**	1.347***
	(0.348)	(0.332)	(0.370)	(0.358)	(0.349)	(0.332)	(0.475)
Social image	0.820**	0.985***	-0.223	-0.012	0.825**	0.967***	1.160**
	(0.368)	(0.357)	(0.396)	(0.382)	(0.369)	(0.354)	(0.516)
Girl (=1)					0.522	0.459	
					(0.369)	(0.349)	
*Self-image**Girl					-0.653	-0.631	
					(0.506)	(0.485)	
*Social image**Girl					-1.047*	-0.922*	
					(0.539)	(0.507)	
Control variables	No	Yes	No	Yes	No	Yes	No
Constant	2.807***	-2.517	3.314***	0.615	2.799***	-1.172	2.307***
	(0.255)	(2.812)	(0.274)	(3.664)	(0.255)	(2.462)	(0.347)
Observations	180	180	186	186	366	366	86
Pseudo R-sq.	0.008	0.033	0.001	0.020	0.004	0.023	0.024
	Boys 7-10y	Boys 11-14y		Girls 7-10y		Girls 11-14y	
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Self-image	1.223**	-0.026	-0.007	0.077	-0.020	-0.013	0.094
	(0.464)	(0.491)	(0.490)	(0.574)	(0.572)	(0.478)	(0.450)
Social image	1.168**	0.457	0.738	-0.220	0.199	-0.243	0.021
	(0.473)	(0.509)	(0.506)	(0.680)	(0.611)	(0.472)	(0.468)
Control variables	Yes	No	Yes	No	Yes	No	Yes
Constant	1.377	3.308***	-1.041	2.920***	-2.572	3.603***	-0.159
	(3.407)	(0.351)	(5.832)	(0.436)	(7.625)	(0.346)	(2.686)
Observations	86	94	94	80	80	106	106
Pseudo R-sq.	0.069	0.003	0.034	0.001	0.021	0.001	0.040

Notes: The table presents tobit regressions using given amount of tokens as the dependent variable with censoring at the amounts of 0 and 10. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Control variables include daily pocket allowance, standardized measure of patience, number of siblings, child's age, cognitive abilities normalized at the age cohort level, and dummy variables for hearing any information about the experiment before the session, attempt of communication from dictator towards observer, attempt of communication from observer towards dictator, and school fixed effects.

Table 4.9. OLS regressions estimates of the effect of observer's gender in *Social image* treatment

Variables	Dependent variable: given amount of tokens			
	Social image treatment			
	Boys		Girls	
	(1)	(2)	(3)	(4)
Girl observer (=1)	-0.437	-0.601	-0.138	0.018
	(0.509)	(0.515)	(0.520)	(0.533)
Controls	No	Yes	No	Yes
Constant	3.923***	-2.686	3.281***	8.345**
	(0.390)	(3.464)	(0.355)	(3.161)
Observations	63	63	60	60
R-squared	0.012	0.200	0.001	0.307

Notes: The table presents OLS regressions using given amount of tokens as the dependent variable. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Control variables include daily pocket allowance, standardized measure of patience, number of siblings, child's age, cognitive abilities normalized at the age cohort level, and dummy variables for hearing any information about the experiment before the session, attempt of communication from dictator towards observer, attempt of communication from observer towards dictator, and school fixed effects.

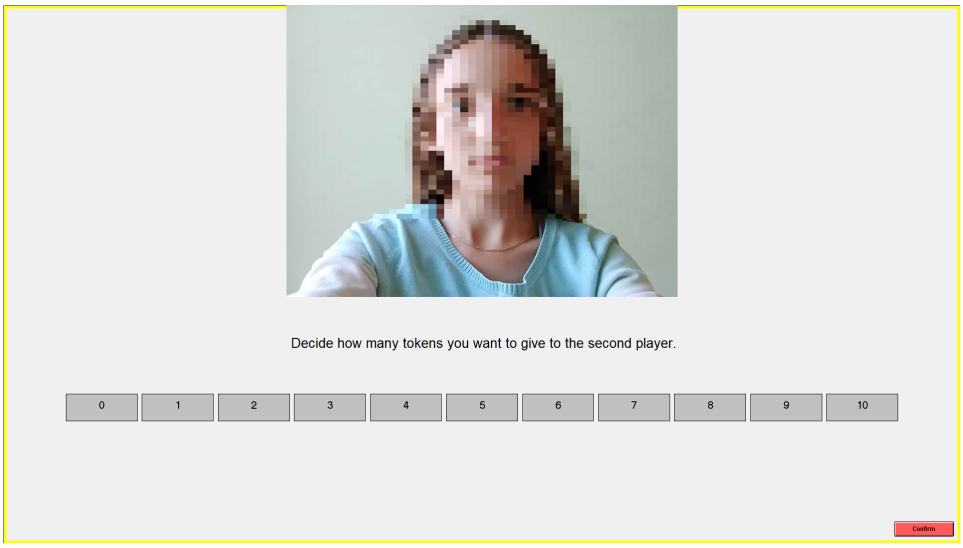

Figure 4.2. Decision screen in *Self-image* and *Social image* treatment

Notes: Translated from Croatian, picture is blurred for privacy reasons.

Figure 4.3. Decision screen in *Control* treatment with a screenshot of the video used with 2nd and 4th-grade subjects

Notes: Translated from Croatian.

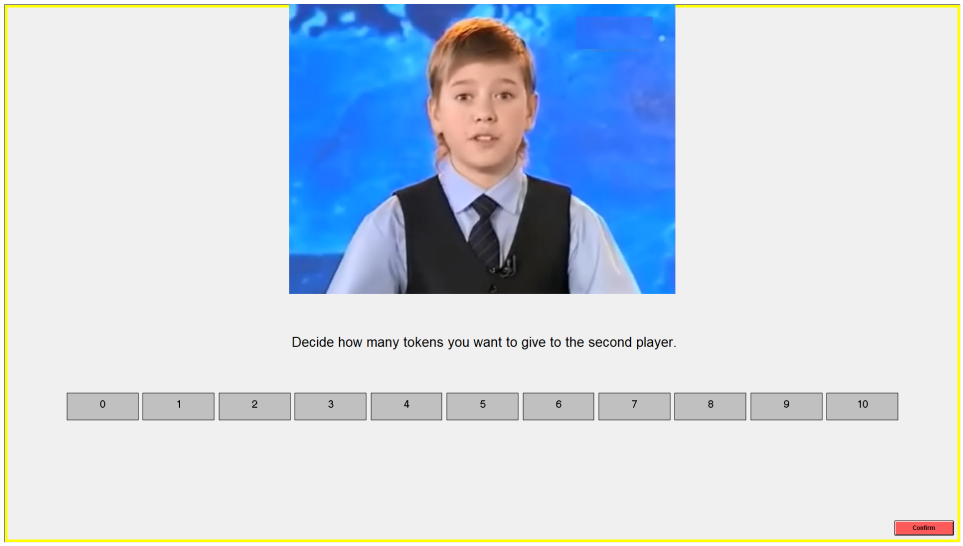

Figure 4.4. Decision screen in *Control* treatment with a screenshot of the video used with 6th and 8th-grade subjects

Notes: Translated from Croatian.

Figure 4.5. Lab in the field during an experimental session

5

The Development of Egalitarian Norm Enforcement in Childhood and Adolescence *

Joint with Armin Falk and Fabian Kosse

5.1 Introduction

Egalitarian norm is a long-existing and widely-pervasive organizing principle. It shapes our everyday behavior, e.g., when we split the bill in a restaurant or divide the profits from a business venture, and it functions as a crucial driver behind prosocial behavior (see Camerer, 2003; Engel, 2011), determining the way in which it is frequently modeled (Fehr and K. M. Schmidt, 1999; Falk and Fischbacher, 2006; Andreoni and Bernheim, 2009). To ensure that people act according to the norm, societies rely on enforcement mechanisms. In particular, if a person violates the egalitarian norm, unaffected third-parties commonly punish such behavior (e.g. Fehr and Fischbacher, 2003, 2004). The extent of punishment might vary, but the punishment itself is robust and widespread, observed in societies around the globe (Henrich et al., 2006; Bernhard et al., 2006). However, while many studies focus on egalitarian norm enforcement among adults, much less is known about its emergence within human development.

* We thank the headmasters of schools and the parents of children for allowing us to perform the research as well as Ksenija Bašić, Matko Bišćan, Ana Kuric, Silvija Lučevnjak, Sanja Portnar and Zrinka Turalija for excellent research support. Moreover, we thank various seminar or workshop participants at the University of California, San Diego, Middlesex University, Ludwig Maximilian University of Munich, University of Mainz, University of Bonn and Kiel Institute for the World Economy for helpful comments. We gratefully acknowledge financial support by Leibniz Programme of the German Science Foundation (DFG).

Despite a rich body of literature investigating the development of children's other-regarding preferences (e.g. Fehr et al., 2008, 2013; Almås et al., 2010; Falk and Kosse, 2016; Deckers et al., 2017; Brocas et al., 2017; Sutter et al., 2018), less focus is devoted to norm-enforcing behavior. Given its relevance, it holds strong interest to shed more light on the roots of egalitarian norm enforcement. In particular, as children become less selfish and more willing to share as they get older (Harbaugh et al., 2003; Benenson et al., 2007; Fehr et al., 2008, 2013; Bauer et al., 2014; Kosse et al., forthcoming), it is essential to understand how enforcement, which underpins such behavior, develops. Therefore, in this paper we investigate the emergence and development of egalitarian norm enforcement. More specifically, we take the most commonly-used third-party punishment game where a third-party is added to a regular dictator game (Fehr and Fischbacher, 2003, 2004; Henrich et al., 2006; Bernhard et al., 2006; Lewisch et al., 2011), adapt it for children's understanding and run an experiment with 635 children and adolescents aged 9-18.

We show that children start enforcing the egalitarian norm at the age of 11-12. In addition, we show that: (i) as the norm enforcement emerges, a non-negligible number of punishers also disapprove of highly generous transfers that exceed the egalitarian norm, (ii) the behavior of punishers only changes until 13-14 years of age, indicating that egalitarian norm enforcement is mainly developed by that period, and (iii) the dictators increase their transfer towards the egalitarian norm primarily in same period when the norm enforcement develops.

Our study contributes to several streams of literature. Primarily, it contributes to the literature on the development of other-regarding behavior, as we show that the egalitarian norm enforcement mainly develops from 11 to 14 years of age. Moreover, we replicate the most common finding in the mentioned literature, observing that children's selfish behavior decreases with age as older dictators give more on average than the younger ones (Harbaugh et al., 2003; Benenson et al., 2007; Fehr et al., 2008, 2013; Bauer et al., 2014; Kosse et al., forthcoming). Interestingly, we find that the decrease of selfishness primarily occurs in the period when the egalitarian norm enforcement develops, suggesting a connection between the two. This provides supporting evidence that punishment is one of the core mechanisms contributing to the commonly-observed increase of prosocial behavior with increasing age.

Most closely to our paper, our results connect to other third-party punishment studies with young children. Lergetporer et al. (2014) find that 7-11 year-old children do not enforce the norm of cooperation, while McAuliffe et al. (2015) and Jordan et al. (2014) find that 6-8 year-old children punish deviations from unfair allocations. Our findings are more in line with the latter, as our youngest cohort also engages in punishment; however, in contrast to McAuliffe et al. (2015) and Jordan et al. (2014), we find that the youngest participants

do not enforce the norm, as we still observe substantial punishment at the equal share. The difference between our findings could be explained by differences in design and through spiteful preferences or preferences towards advantageous inequality, which can still be highly present in this development period (Fehr et al. (2013) and McAuliffe et al. (2014); see discussion in Section 5.4). Complementing the aforementioned studies with young children, our study connects the behavior from childhood to adulthood, as we record punishment from 9 to 18 years of age. We find that enforcement behavior mainly develops by mid-adolescence. This supports the major contribution of Almås et al. (2010), who observe that children's understanding of fairness still strongly evolves in the mentioned period. Furthermore, it is in line with the findings of Gummerum and Chu (2014), who study the development of concerns for outcomes and intentions in second and third parties. While the authors do not observe a change in third-parties' concerns for outcomes or intentions in this period, they report an increase in punishment behavior when moving from pre-adolescence to mid-adolescence. Looking at our results with older adolescents, we observe that the punishment behavior does not change from mid-adolescence to adulthood, where our study connects to other third-party punishment studies with adults (Fehr and Fischbacher, 2003, 2004; Henrich et al., 2006; Bernhard et al., 2006; Lewisch et al., 2011). More specifically, the behavior of the oldest cohorts in our study is comparable to studies using a similar design with adults (Bernhard et al., 2006).[1]

Finally, our results indicate that a non-negligible proportion of children approve of the egalitarian norm but not above-egalitarian transfers. The finding relates to other third-party punishment studies (Henrich et al., 2006; Gummerum and Chu, 2014) and complements the growing literature on antisocial punishment (Herrmann et al., 2008; Parks and Stone, 2010; Irwin and Horne, 2013). Importantly, it connects directly to studies focusing on norm elicitation, where it has been shown that in contrast to the egalitarian transfer in a dictator game, a substantial proportion of subjects do not approve of above-egalitarian transfers (Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016). Our findings suggest that this normative valuation stretches its roots to a much younger age. As soon as the norm enforcement emerges, not only selfish but also generous deviations can be viewed as less appropriate.

The remainder of the paper is structured as follows. In Section 5.2, we explain the experimental design and procedure. In Section 5.3, we report the results of our study. In Section 5.4, we discuss the potential mechanisms behind our findings, before we finally conclude the paper in Section 5.5.

[1] While many studies implement the third-party punishment game where a punisher is added to a regular dictator game, we limit the punisher's action space in the same way as in Bernhard et al. (2006).

5.2 Experimental design and procedure

We divide the following section into two subsections. In the first subsection, we explain the experimental design. In the second subsection, we describe the experimental procedure and the subject pool.

5.2.1 Experimental design

The experimental game we used comprises three players: a dictator, recipient and punisher (see Figure 5.1). At the beginning of the game, the dictator is endowed with 10 tokens and has to divide them between him- or herself and the recipient, who has no endowment. The punisher is endowed with 5 tokens. After the dictator makes his or her decision, the punisher can punish the dictator with up to 2 tokens. For each token that the punisher uses, the dictator's payoff decreases by 3 but is bounded by a minimum of 0. We limit the choice of punishment to a maximum of 2 tokens to ensure comprehension with participants of all age. To enforce the egalitarian norm, a subject will not punish on the transfer of 5 but will punish on selfish transfers that deviate from 5 (Fehr and Fischbacher, 2003, 2004; Henrich et al., 2006; Bernhard et al., 2006).

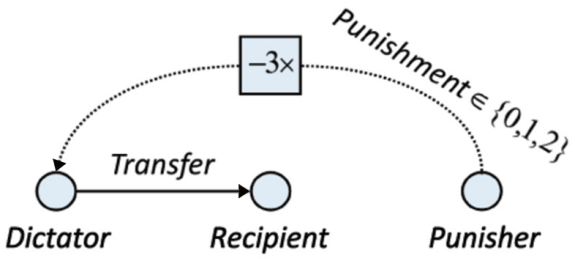

Figure 5.1. Dictator game with a third-party punisher

Notes: Dictator is endowed with 10 tokens and has to divide them between him- or herself and the recipient, who has no endowment. The punisher is endowed with 5 tokens. After the dictator's decision, the punisher can punish the dictator with up to 2 of his or her tokens. For each token that the punisher uses, the dictator's payoff decreases by 3 but is bounded by a minimum of 0.

To gain a comprehensive picture, we implemented the strategy method. Each punisher had to indicate how much he or she would punish conditional on every

potential choice of the dictator.[2] We devoted special focus to its implementation. Instead of a list, the punishers were presented with 11 ordered pages. On each page, there was one potential transfer from the dictator with depicted punishment options and vividly-displayed outcomes for each of those options (see Figure 5.5 in Appendix). This ensured that subjects of all ages and different cognitive abilities were fully aware of the potential payoff outcomes conditional on their choices.

5.2.2 Procedure and subject pool

The study was conducted in Našice, a small city in eastern Croatia with an approximate population of 16,000. The participants were students in 3rd, 5th and 7th grade of elementary school and 1st and 3rd grade of high school. The age of the children was 9-10, 11-12, 13-14, 15-16 and 17-18, respectively.[3] Altogether, 635 students from two elementary schools and one high school participated in the study. The experiment was conducted during regular school hours. Sessions were organized for each class and each student was assigned one role in the game, which was the same for everybody in a class. Our approach allowed us to ensure the avoidance of selection bias when comparing high school and elementary school subjects. In particular, the participating schools were the only schools in the local area. Moreover, at the time of the experiment, the participating high school had the largest number of students among high schools in Croatia, and it offered a variety of study programs to its students. This allowed us to balance the sample across the different types of programs within each high school age cohort (see Table 5.1).[4] Three participating classes of students were from three-year high school programs, and students from different classes were assigned a different role in the game. Three-year programs are considered the least difficult to finish and have the lowest enrollment requirements. Furthermore, three participating classes of students were from four-year non-gymnasium programs and three were from four-year gymnasium programs. The latter are considered the most difficult programs and have the highest enrollment requirements. For both gymnasium and non-gymnasium four-year programs, students from different classes were also assigned different roles in the game. For elementary

[2] Note that the strategy method was shown not to influence results of third-party punishment (Jordan et al., 2016), and was successfully used with 7 year-old children (Lergetporer et al., 2014).

[3] Our sample contains few exceptions where a child is older or younger than his or her classmates, e.g., if a child fails a grade or starts school earlier. In our main comparison, we follow the usual approach in the literature and bin the children based on their age in their corresponding age cohorts (9-10, 11-12, 13-14, 15-16 and 17-18 years of age). Classifying the children strictly according to their grade does not change our results.

[4] Unlike elementary school education in Croatia, high school education offers multiple study programs, which students choose based on their preference and prior academic achievement.

schools, three classes were participating from each grade level and school, and children across these classes were assigned different roles in the game.[5]

Table 5.1. Sampling

Grade (age)	Elementary school 1	Elementary school 2	High school		
			3 year program	4 year non-gymn. program	4 year gymn. program
3rd (9-10)	3 classes	4 classes			
5th (11-12)	3 classes	3 classes			
7th (13-14)	3 classes	4 classes			
1st (15-16)			3 classes	3 classes	3 classes
3rd (17-18)			3 classes	3 classes	3 classes

Notes: The table presents an overview of the amount of participating classes per school, program and grade level.

During and after the experiment, the players remained anonymous to each other, although they were aware that the other players were students of similar age but not from their class. Each student was given a code that served as identification during the experimental procedure and no names were used. During the experiment, only the experimenters and the children were allowed to be in the classroom. The environment during the experiment resembled the environment during a normal school test. The tables where children were sitting were separated and children were not allowed to talk to each other (see Figure 5.6 in Appendix). The experimental instructions comprised oral instructions read by the same experimenter in all experimental sessions (for instructions, see Appendix 5.B). The oral instructions were supported with posters on the blackboard, which were incorporated in the structured explanation of the game. The use of posters with displayed steps of the game was an approach specifically aimed at assisting in the children's understanding of the game. To ensure

[5] In two grade levels in elementary school 2, we added an additional class to have a balanced sample size across the two schools. In these two cases, children from the extra class were assigned the same role in the game as in one of the other three classes.

standardization, alongside the main explanation of the game, the posters also contained several examples that were thoroughly explained for each session.

Before the start of the game, each student had to answer three control questions, each comprising three sub-questions. If a child had problems solving any of the control questions, an experimenter explained the rules of the game again just to that child, and assisted in answering the first control question through explanation. The second and third questions were used as a control of understanding. Out of 217 punishers, 180 answered the control questions correctly, while out of 209 dictators, 193 managed the same. The children who did not answer the control questions correctly were excluded from the analysis.[6] As the number of participating children across classrooms did not perfectly match, 8 dictators were each matched with 2 punishers, and 2 recipients were each matched with 2 dictators. After the experimental game, subjects answered a short questionnaire and took a cognitive abilities test comprising 16 matrices items.

The experiment was approved by the school principals of all three schools. Parents of participating children gave written consent for participation after they were informed about the nature and possible consequences of the study. With the consent, parents also filled out a small questionnaire with demographic and socio-economic questions. The tokens that students earned in the game were converted to Croatian kunas for payment, and were paid out within 2 weeks after the experiment. The value of a token was adjusted to the age cohort, whereby students from 3rd grade of elementary school earned 3 kunas for a token, students from 5th and 7th grade of elementary school earned 4 kunas for a token, and students from 1st and 3rd grade of high school earned 5 kunas for a token, where 1 HRK $\approx$ 0.17 USD at the time of the experiment. In addition to the earnings in the game, each subject also received a participation fee equivalent to the value of 2 tokens.

5.3 Results

In the following, we first analyze the behavior of the punishers in two steps: first, we apply simple mean comparisons to identify when subjects start enforcing the egalitarian norm; and second, we construct a piecewise linear regression model to analyze the entire punishment patterns across different age cohorts. After

[6] Including these participants into the analysis yields similar results. Specifically, the punishment patterns stay very similar across all age cohorts, while the dictators exhibit an almost identical development pattern of transfers across age. In addition to those that did not answer the control questions correctly, we excluded two children with mental disabilities from our data. Nevertheless, they participated, and given that they were unable to finish the tasks, they were paid the mean earning of their age cohort.

analyzing the punishers' behavior, we analyze the development of the dictators' behavior.

5.3.1 Behavior of the punishers

As a first step of our analysis of punishment behavior, we explore when subjects start enforcing the egalitarian norm. For this purpose, we first compare punishment for egalitarian transfer and selfish deviations from the transfer. This comparison is contrasted through age cohort averages in Figure 5.2. The left bar represents average punishment in response to selfish transfers (i.e., transfer < 5) and the middle bar represents punishment in response to egalitarian transfer (i.e., transfer $= 5$). The average punishment for selfish transfers is approximately 1 token for all age cohorts. For the youngest cohort, the egalitarian transfer does not provoke a significantly different punishment than selfish transfers ($p = 0.147$, two-sided t-test, N = 32). For age cohorts 11-12 and older, the average punishment on egalitarian transfer decreases and is significantly smaller compared to punishment for selfish transfers ($p < 0.001$ for each age cohort above 9-10, two-sided t-test, N(11-12) = 25, N(13-14) = 31, N(15-16) = 39, N(17-18) = 53). The result is driven by the fact that the proportion of children punishing the egalitarian transfer sharply decreases for the older cohorts. While for the 9-10 age cohort 65% of children punish on egalitarian transfer, the amount decreases to, e.g., 32% for the 11-12 cohort. This indicates that starting at age 11-12, the majority of children approve of the egalitarian norm and selfish deviations from the norm provoke higher punishment. A robustness check controlling for individual differences in, e.g., gender and parental income supports this finding (see Appendix 5.A).

Next, we analyze punishers' behavior on generous transfers. In Figure 5.2, with the right bar, we also present punishment in response to a marginal generous deviation from the norm (i.e., transfer $= 6$). Note that we only focus on punishment in response to a transfer of 6. As the transfer increases beyond 6, the punishment decreases, which is mechanically driven (see Figure 5.3; given very high transfers, for the punisher using only one punishment token is sufficient to make the dictator end with zero tokens). We observe that for the youngest cohort the punishment for marginal generous deviation is slightly smaller than for egalitarian transfer. Together, the bars present a decreasing punishment pattern with relatively high punishment for the egalitarian transfer, inconsistent with egalitarian norm enforcement. Beginning at 11-12 years of age, children exhibit lower punishment for egalitarian transfer than for marginal generous deviation from it. Even though the difference is statistically insignificant for most of the cohorts ($p > 0.240$ for age cohorts 11-12, 15-16, 17-18, $p = 0.018$ for age cohort 13-14, two-sided t-test, N(11-12) = 25, N(13-14) = 31, N(15-16) = 39, N(17-18) = 53), it implies that some children do not approve of generous deviations

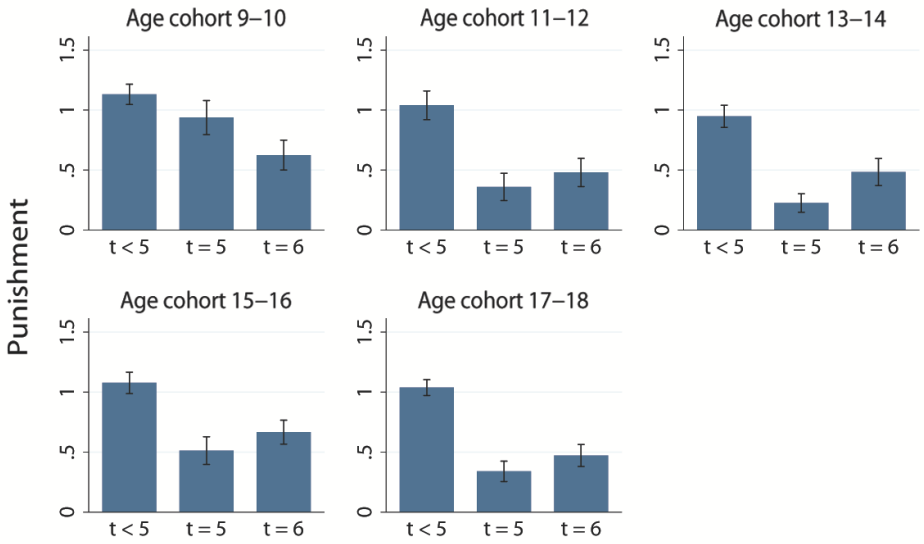

Figure 5.2. Average punishment amount by age cohort

Notes: Average punishment on egalitarian transfer of 5 tokens (t = 5), average of mean punishment on transfers smaller than 5 tokens, i.e., selfish deviations (t < 5) and average punishment on transfers of 6 tokens, i.e., marginal generous deviation (t = 6). Error bars show standard error of the means. Children from age cohorts 11-12 and higher exhibit a highly significant difference between punishment on t = 5 and t < 5 (p = 0.147 for 9-10 age cohort, p < 0.001 for each age cohort above 9-10, two-sided t-test, N(9-10) = 32, N(11-12) = 25, N(13-14) = 31, N(15-16) = 39, N(17-18) = 53).

from the norm and are willing to punish them. In particular, a non-negligible proportion of 11-12 year-olds (20%) do not punish on egalitarian transfer yet punish on marginal generous deviation from it. This finding is robust for all older cohorts (9% for age cohort 9-10, 23-36% for age cohorts above 11-12). We discuss the motivation behind punishment on above-egalitarian transfers as well as the motivation behind the punishment behavior of the youngest cohort in the Discussion section (5.4). Altogether, starting from 11-12 years of age, we observe that the majority of children approve of the egalitarian transfer. On average, selfish deviations, and by some even generous deviations, incur higher punishment. Hence, first indications of the egalitarian norm enforcement occur among 11-12 year-old children.

As a second step of our analysis concerning punishment behavior, we inspect the entire punishment pattern. Figure 5.3 shows the (i) average punishment amount conditional on the dictator's transfer for all age cohorts, and (ii) a plotted piecewise linear regression model that we designed to investigate further development in the punishment pattern starting from 11-12 years of age

(model estimation in Table 5.2). In particular, while our first part of the analysis indicates when subjects start enforcing the egalitarian norm, the model allows us to observe whether there is any further development of punishment after enforcement emerges. The main characteristic of the model is that it allows for two break points (jumps). More specifically, we observe that the main punishment changes happen in response to transfers of 4, 5 and 6; hence, we build the model to detect a potential change in the punishment pattern when contrasting punishment decisions on deviations to 4 and 6 tokens with punishment decisions on larger deviations. The model also allows for different slopes before and after the egalitarian norm (see Equation 1).

The dependent variable y_{it} represents the amount of punishment chosen by a third party i in reaction to a dictator's transfer of $t \in \{0, 1, ..., 10\}$, where t is used to estimate the change in punishment for transfers 0 to 4. d_{1t} is a dummy variable used to estimate the change (jump) in the punishment pattern when going from a transfer of 4 to 5, whereby it is 0 for $t < 5$ and 1 for $t \geq 5$. d_{2t} is a dummy variable used to estimate the change (jump) in the punishment pattern when going from a transfer of 5 to 6, whereby it is 0 for $t < 6$ and 1 for $t \geq 6$. t' allows for different slopes before and after the egalitarian norm, hence, it is 0 for $t < 5$ and $t - 5$ for $t \geq 5$.[7]

$$y_{it} = \beta_0 + \beta_1 t + \beta_2 d_{1t} + \beta_3 d_{2t} + \beta_4 t' + \epsilon_{it} \quad (1)$$

First, we inspect the average punishment patterns as presented in Figure 5.3. 9-10 year-old children exhibit a downward-sloping punishment pattern, where the response to egalitarian transfer does not stand out from the rest. Starting with 11-12 years of age, the egalitarian transfer incurs much less punishment and on average becomes the least-punished transfer (not including highest transfers). This confirms our results from the first part of the analysis.

Next, we inspect the development of the punishment pattern starting from the 11-12 age cohort using the estimated coefficients of our regression model (see Table 5.2). For 11-12 year-old children, the pattern indicates that a marginal selfish deviation from the egalitarian norm is directly answered by pronounced punishment in comparison to greater deviations (Column (2), coefficient of the punishment jump for transfers 4 to 5 (β_2), $p = 0.033$, two-sided t-test, N = 275). The punishment for greater selfish deviations increases gradually (slope regarding transfers 0 to 4 (β_1), $p < 0.001$, two-sided t-test, N = 275). In contrast to 11-12 year-olds, 13-14 year-old children exhibit an approximately two

[7] We apply clustered standard errors, which takes into account the fact that punishment decisions are dependent within the subject but independent across subjects.

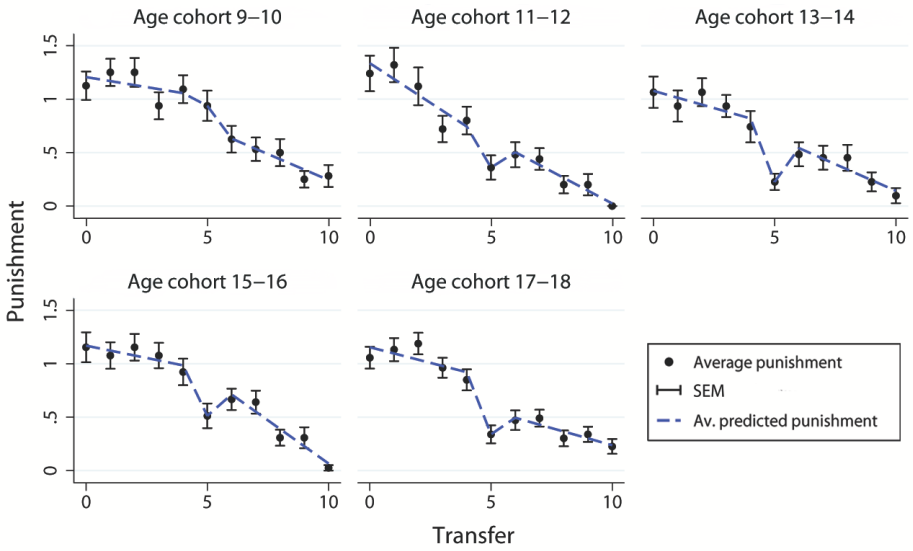

Figure 5.3. Actual and predicted punishment for all transfers by age cohort

Notes: The black dots show the average observed punishment. The blue dashed line shows the prediction of punishment based on our piecewise linear model, which allows for a jump in punishment on marginal deviations from the egalitarian norm and different slopes before and after the egalitarian norm. Error bars show standard errors of the means.

times larger punishment response for the marginal selfish deviation in comparison to greater deviations (Column (3), coefficient of the punishment jump for transfers 4 to 5 (β_2), $p = 0.001$, two-sided t-test, N = 341). Simultaneously, there is a smaller increase for greater selfish deviations than among 11-12 year-old children (slope regarding transfers 0 to 4 (β_1), $p = 0.239$, two-sided t-test, N = 341). This punishment pattern of pronounced response to the marginal selfish deviation and a small punishment increase on greater deviations continues for 15-16 and 17-18 year-old children (Columns (4) and (5)). Together, the regressions show that although the egalitarian norm enforcement is first visible at 11-12 years of age, the punishment pattern still develops until 13-14 years of age. As the subjects approach adulthood, we do not identify any further developments of the pattern.

Table 5.2. Piecewise linear regression models

Variables	Dependent variable: punishment amount				
Age cohort:	9-10	11-12	13-14	15-16	17-18
	(1)	(2)	(3)	(4)	(5)
Transfer (β_1)	-0.038	-0.148***	-0.065	-0.046	-0.058**
	(0.039)	(0.033)	(0.054)	(0.035)	(0.028)
Dummy 4-5 (β_2)	-0.081	-0.236**	-0.529***	-0.426***	-0.523***
	(0.178)	(0.105)	(0.151)	(0.140)	(0.114)
Dummy 5-6 (β_3)	-0.209	0.264*	0.416***	0.362**	0.219*
	(0.144)	(0.151)	(0.136)	(0.172)	(0.111)
Transfer' (β_4)	-0.059	0.028	-0.035	-0.115**	-0.006
	(0.059)	(0.037)	(0.081)	(0.050)	(0.042)
Constant (β_0)	1.206***	1.336***	1.077***	1.169***	1.155***
	(0.108)	(0.164)	(0.149)	(0.124)	(0.089)
Observations	352	275	341	429	583
R-squared	0.204	0.319	0.211	0.232	0.222

Notes: The table presents OLS regression models using punishment as the dependent variable. Clustered standard errors at the individual level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. β_1 estimates the slope for transfers 0 to 4 (Transfer (t) $\in$ 0, 1, ..., 10). β_2 estimates the change (jump) in the punishment pattern when going from a transfer of 4 to 5 (Dummy 4-5 = 0 for t < 5 and 1 for $t \geq 5$). β_3 estimates the change (jump) in the punishment pattern when going from a transfer of 5 to 6 (Dummy 5-6 = 0 for t < 6 and 1 for $t \geq 6$). β_4 estimates the change in the slope for $t \geq 6$ (Transfer' = 0 for t < 5 and t– 5 for $t \geq 5$).

On the generous side of the egalitarian norm, we observe that starting from 11-12 years of age, deviations from the norm are also answered by an increase in punishment (coefficient of the punishment jump for transfers 5 to 6 (β_3), $p = 0.094$, two-sided t-test, N = 275). The behavior does not exhibit any developmental pattern from 11-12 to 17-18 year-olds, but it further supports our prior observation regarding punishment for marginal generous deviation from the norm.

5.3.2 Behavior of the dictators

In the last step of our analysis, we inspect the sharing behavior of the dictators. As the punishers' behavior reveals the emergence of the egalitarian norm enforcement, we investigate whether this is reflected in the dictators' behav-

ior. Figure 5.4 shows the average transfers by age cohort. Children from the youngest cohort transfer the smallest average amount (1.86 tokens). Regressing the amount of transfer on the age cohort yields a positive coefficient ($p = 0.039$, OLS regression model, N = 193). However, this increase in transfers seems to be non-linear. It primarily occurs from 11-12 to 13-14 years of age, where the average transfer increases from 1.91 to 3.35 ($p = 0.016$, two-sided t-test, N(11-12) = 33, N(13-14) = 33). For the older cohorts, the transfer decreases slightly in comparison to the 13-14 age cohort, although the decrease is not statistically significant ($p > 0.179$ for both 15-16 and 17-18 age cohorts, two-sided t-test, N(13-14) = 33, N(15-16) = 61, N(17-18)= 44). This pattern indicates that the development of the dictators' behavior mirrors that of the punishers' behavior. Specifically, we observe that the egalitarian norm enforcement mainly develops from 11 to 14 years of age. Simultaneously, dictators change their behavior in the direction of the egalitarian norm in the corresponding time period.

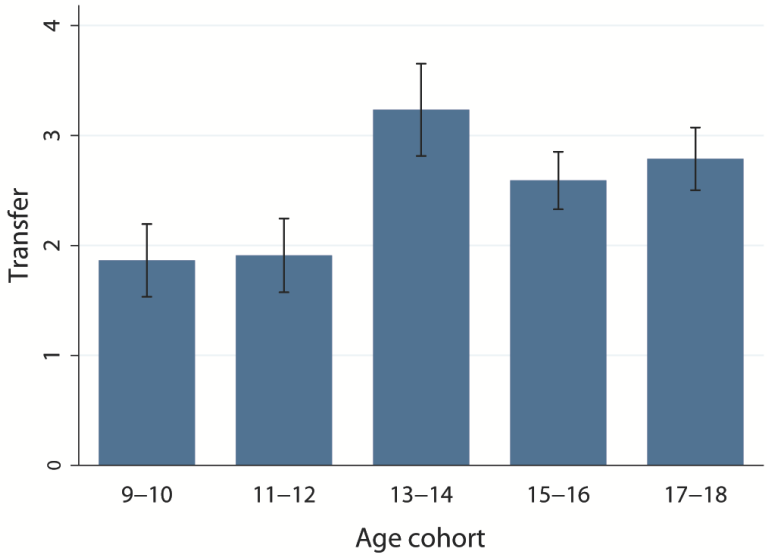

Figure 5.4. Average dictator's transfer by age cohort

Notes: Error bars show standard errors of the means. There is an upward trend of transfers by age cohort. Regressing the amount of transfer on the age cohort yields a positive coefficient (p = 0.039, OLS regression model, N = 193); however, the increase seems to be non-linear. It primarily occurs from 11-12 to 13-14 years of age, where the average transfer increases from 1.91 to 3.35 (p = 0.016, two-sided t-test, N(11-12) = 33, N(13-14) = 33).

5.4 Discussion

In this section, we discuss potential mechanisms behind our findings. First, we focus on the punishment behavior of our youngest cohort: The 9-10 year-old children use the punishment tool but not to enforce the egalitarian norm, as they gradually decrease their punishment with the dictator's transfer and still punish the equal share. These findings relate to the discussion on the development of other-regarding behavior. On the one hand, our findings are at odds with studies reporting prevalent egalitarian behavior at the age of our youngest cohort. Most popularly, Fehr et al. (2008) use three simple games to classify 3-8 year-old children in other-regarding types. They observe that the majority of children (60%) become egalitarian by the age of 7-8. They find that other preferences, such as spiteful preferences where the subject's utility decreases with the other's payoff, diminish by that age. On the other hand, several studies report that spiteful preferences persist longer, offering a potential explanation for our findings. Fehr et al. (2013) also classify other-regarding types using a more similar experimental procedure to ours[8], finding that the majority of subjects are still spiteful (42%) at the age of 8-9 (alternatively, these subjects could also care about advantageous inequality, i.e., they prefer to have more than others). Moreover, McAuliffe et al. (2014) also find high levels of spiteful behavior for similarly-aged children as our youngest cohort, while both McAuliffe et al. (2014) and Fehr et al. (2013) report that this type of behavior decreases with age. Hence, it seems unclear when spitefulness exactly diminishes, and in line with the latter studies, our youngest punishers could have spiteful preferences and punish the dictators simply to reduce their payoff, or they could also care about advantageous inequality and punish the dictators to have more than them.

Furthermore, the comparison with Jordan et al. (2014) and McAuliffe et al. (2015) also supports this explanation. In particular, McAuliffe et al. (2015)[9] run an experiment with 5-6 year-old children in which a third party can punish unfair allocations between two players. In comparison to our design, the punisher always obtains a much larger endowment in the one-shot game; hence, the punisher is not comparing him- or herself with others in the same egalitarian way as in our design. In contrast to the findings with our youngest cohort, they observe that 6 year-old children already do not punish much on the fair choice. This difference could be explained by spiteful preferences, or preferences towards

[8] While both Fehr et al. (2008) and Fehr et al. (2013) use the same binary games, in Fehr et al. (2013) (i) the experiment was conducted in large group sessions in a classroom and not one-to-one with the experimenter, and (ii) the children were incentivized with money rather than candy. The same holds true for our experiment.

[9] Jordan et al. (2014) run an experiment with the same game and similarly-aged children as in McAuliffe et al. (2015), but they investigate the influence of in- and out-group bias and its emergence.

advantageous inequality.[10] In support of this explanation, the findings of Gummerum and Chu (2014) are more fitting with the behavior of our youngest cohort. While our design has several differences compared with that of Gummerum and Chu, the punisher's endowment is not one of them.[11] In particular, the authors investigate the development of concerns for outcomes and intentions with second- and third-party punishers. In order to manipulate intentions, they run several mini-third-party punishment games, where the dictator can only choose between two division options. If we look at the specific game where dictators choose between dividing equally or keeping most for themselves, the authors report a substantial average amount spent on punishing the equal share for 8 year-olds, which decreases for older subjects. This is in line with our findings.

Conclusively, while we cannot deduce a single motive, spiteful preferences or preferences towards advantageous inequality are a plausible explanation for the behavior of our youngest cohort.[12] Moreover, the comparison between our study and that of McAuliffe et al. (2015) might offer a valuable insight. McAuliffe et al. indicate that children already care about the fair allocation between two subjects at a very young age. However, in an egalitarian setup where all subjects possess the same endowment, our study would suggest that children are still driven by other factors at the age of 9-10, i.e., they might still be driven by self-comparison motives such as spiteful preferences or preferences towards advantageous inequality. Hence, if children's understanding of a normative violation exists at a younger age, it might be dominated by other factors until 11-12 years of age.

Next, we discuss our finding regarding above-egalitarian punishment. We observe that a non-negligible proportion of children approve of the egalitarian but not above-egalitarian transfers. This behavior is inconsistent with outcome-based models of other-regarding behavior (e.g. Fehr and K. M. Schmidt, 1999), but could be explained with social norm based models (Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016). While both types of models emphasize the normative value of the equal share, the latter posit that each action has a social-normative value that indicates its appropriateness. Furthermore, Krupka and

[10] In the case of using spitefulness as an explanation, a subject's utility component representing the influence of other's payoff should be concave in the payoff differences (at least in the advantageous domain), i.e., reducing the other player's payoff by 1 yields a larger utility increase if payoffs are close than if the other player has a much smaller payoff.

[11] Gummerum and Chu (2014) implement several mini-third-party games, where the dictator always has two possibilities how to divide 10 points. He or she can choose to give 2 points or an alternative allocation. Depending on the game, the alternative allocation can be 0, 2, 5 or 8 points. The punisher is endowed with 5 points, and can punish the dictator with up to 5 points. Each points reduces the dictator's payoff by 2.

[12] Another explanation for the punishment pattern that we observe is that young children only approve of overly generous transfers. However, this explanation lacks support in the literature and might also seem unintuitive.

Weber (2013) and Kimbrough and Vostroknutov (2016) show through norm elicitation methods that in the dictator game people on average approve more of the egalitarian transfer than above-egalitarian transfers. More specifically, there is a consensus about the appropriateness of the egalitarian transfer, yet less so with above-egalitarian transfers. While some subjects approve of transferring high amounts, others do not. This is in line with our observation that only some subjects engage in punishment of above-egalitarian behavior. Alternatively, one can explain these results through descriptive norms, i.e., each deviation from the usual behavior encounters punishment (Irwin and Horne, 2013), or one can think of the above-egalitarian punishment as “derogation of the do-gooders" (Monin, 2007).

Finally, we discuss the finding that dictators’ transfers primarily increase in the period as the norm enforcement develops, suggesting a connection between the two. This provides supporting evidence that punishment is one of the core mechanisms contributing to the commonly-observed increase of prosocial behavior with age (Harbaugh et al., 2003; Benenson et al., 2007; Fehr et al., 2008, 2013; Bauer et al., 2014; Kosse et al., forthcoming). Indeed, the mere threat of punishment is crucial for the prosocial and cooperative behavior of adults (Fehr and Gächter, 2002; Herrmann et al., 2008). Alternatively, it is possible that the observed increase of dictators’ transfers is not due to the emergence of punishment. An underlying variable, e.g., the internalization of norms could be driving the development both norm-enforcing punishment and prosocial behavior as we observe it. While investigating the internalization of norms is a very interesting question, it remains rather intangible to actually confirm when a behavior starts being a norm. Nevertheless, several studies indicate an understanding of norms at a very young age (e.g. Fehr et al., 2008; M. F. H. Schmidt and Sommerville, 2011), which questions the argument that the internalization of norms could be solely responsible for the observed simultaneous changes in behavior during the period of 11-14 years of age.

5.5 Conclusion

In this paper, we have investigated the emergence and development of egalitarian norm enforcement in childhood and adolescence. We took the most commonly-used third-party punishment game for adults, adapted it to ensure children’s understanding and ran an experiment with 635 subjects aged 9-18. We show that enforcement of the egalitarian norm starts at the age of 11-12, and mainly develops until the age of 13-14. Furthermore, we find that as the egalitarian norm enforcement emerges, a non-negligible proportion of punishers also disapprove of transfers that go above the egalitarian norm. Finally, we observe that the dictators increase their average transfer towards the egalitarian

norm primarily in the period when we observe that norm enforcement develops. Our results suggest punishment as one of the core mechanisms that underpins children's increased prosocial behavior with increasing age. Furthermore, our findings suggest that the negative valuation of exceeding the norm, which was reported in studies with adults, has roots in a much younger age.

References

Almås, Ingvild, Alexander W. Cappelen, Erik Ø. Sørensen, and Bertil Tungodden (2010): "Fairness and the development of inequality acceptance." *Science*, 328, 1176–1178. [96, 97]

Andreoni, James and B. Douglas Bernheim (2009): "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects." *Econometrica*, 77 (5), 1607–1636. [95]

Bauer, Michal, Julie Chytilová, and Barbara Pertold-Gebicka (2014): "Parental background and other-regarding preferences in children." *Experimental Economics*, 17 (1), 24–46. [96, 110]

Benenson, Joyce F., Joanna Pascoe, and Nicola Radmore (2007): "Children's altruistic behavior in the dictator game." *Evolution and Human Behavior*, 28 (3), 168–175. [96, 110]

Bernhard, Helen., Urs Fischbacher, and Ernst Fehr (2006): "Parochial altruism in humans." *Nature*, 442 (7105), 912–915. [95–98]

Brocas, Isabelle, Juan D. Carrillo, and Niree Kodaverdian (2017): "Altruism and strategic giving in children and adolescents." *CEPR Discussion Paper* (DP12288). [96]

Camerer, C. (2003): *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press. [95]

Deckers, Thomas, Armin Falk, Fabian Kosse, Pia Pinger, and Hannah Schildberg-Hörisch (2017): "Socio-economic status and inequalities in children's IQ and economic preferences." *IZA Discussion Paper* (11158). [96]

Engel, Christoph (2011): "Dictator games: A meta study." *Experimental Economics*, 14 (4), 583–610. [95]

Falk, Armin and Urs Fischbacher (2006): "A theory of reciprocity." *Games and Economic Behavior*, 54 (2), 293–315. [95]

Falk, Armin and Fabian Kosse (2016): "Early childhood environment, breastfeeding and the formation of preferences." *HCEO Working Paper* (2016-036). [96]

Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach (2008): "Egalitarianism in young children." *Nature*, 454, 1079–1083. [96, 108, 110]

Fehr, Ernst and Urs Fischbacher (2003): "The nature of human altruism." *Nature*, 425, 785–791. [95–98]

Fehr, Ernst and Urs Fischbacher (2004): "Third-party punishment and social norms." *Evolution and Human Behavior*, 25 (2), 63–87. [95–98]

Fehr, Ernst and Simon Gächter (2002): "Altruistic punishment in humans." *Nature*, 415 (6868), 137–140. [110]

Fehr, Ernst, Daniela Glätzle-Rützler, and Matthias Sutter (2013): "The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence." *European Economic Review*, 64, 369–383. [96, 97, 108, 110]

Fehr, Ernst and Klaus M. Schmidt (1999): "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114 (3), 817–868. [95, 109]

Gummerum, Michaela and Maria T. Chu (2014): "Outcomes and intentions in children's, adolescents', and adults' second-and third-party punishment behavior." *Cognition*, 133 (1), 97–103. [97, 109]

Harbaugh, William T., Kate Krause, and Steven G. Liday Jr. (2003): "Bargaining by children." *University of Oregon Economics Working Paper* (2002-4). [96, 110]

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Esminger, Clark Barrett, Alexander Bolyanatz, Juan C. Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolin Lesorogol, Frank Marlowe, David Tracer, and John Ziker (2006): "Costly punishment across human societies." *Science*, 312, 1767–1770. [95–98]

Herrmann, Benedikt, Christian Thöni, and Simon Gächter (2008): "Antisocial punishment across societies." *Science*, 319, 1362–1367. [97, 110]

Irwin, Kyle and Christine Horne (2013): "A normative explanation of antisocial punishment." *Social Science Research*, 42 (2), 562–570. [97, 110]

Jordan, Jillian J., Katherine McAuliffe, and David Rand (2016): "The effects of endowment size and strategy method on third party punishment." *Experimental Economics*, 19 (4), 741–763. [99]

Jordan, Jillian J., Katherine McAuliffe, and Felix Warneken (2014): "Development of in-group favoritism in children's third-party punishment of selfishness." *Proceedings of the National Academy of Sciences*, 111 (35), 12710–12715. [96, 108]

Kimbrough, Erik O. and Alexander Vostroknutov (2016): "Norms make preferences social." *Journal of the European Economic Association*, 14 (3), 608–638. [97, 109, 110]

Kosse, Fabian, Thomas Deckers, Pia R. Pinger, Hannah Schildberg-Hörisch, and Falk Armin (forthcoming): "The formation of prosociality: Causal evidence on the role of social environment." *Journal of Political Economy*. [96, 110]

Krupka, Erin L. and Roberto A. Weber (2013): "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association*, 11 (3), 495–524. [97, 109]

Lergetporer, Philipp, Silvia Angerer, Daniela Glätzle-Rützler, and Matthias Sutter (2014): "Third-party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation." *Proceedings of the National Academy of Sciences*, 111 (19), 6916–6921. [96, 99]

Lewisch, Peter G., Stefania Ottone, and Ferruccio Ponzano (2011): "Free-riding on altruistic punishment? An experimental comparison of third-party punishment in a stand-alone and in an in-group environment." *Review of Law and Economics*, 7 (1), 161–190. [96, 97]

McAuliffe, Katherine, Peter R. Blake, and Felix Warneken (2014): "Children reject inequity out of spite." *Biology Letters*, 10 (12), 20140743. [97, 108]

McAuliffe, Katherine, Jillian J. Jordan, and Felix Warneken (2015): "Costly third-party punishment in young children." *Cognition*, 134, 1–10. [96, 108, 109]

Monin, Benoît (2007): "Holier than me? Threatening social comparison in the moral domain." *Revue Internationale de Psychologie Sociale*, 20 (1), 53–68. [110]

Parks, Craig D. and Asako B. Stone (2010): "The desire to expel unselfish members from the group." *Journal of Personality and Social Psychology*, 99 (2), 303–310. [97]

Schmidt, Marco F. H. and Jessica A. Sommerville (2011): "Fairness expectations and altruistic sharing in 15-month-old human infants." *PloS one*, 6 (10), e23223. [110]

Sutter, Matthias, Francesco Feri, Daniela Glätzle-Rützler, Martin G. Kocher, Peter Martinsson, and Katarina Nordblom (2018): "Social preferences in childhood and adolescence. A large-scale experiment to estimate primary and secondary motivations." *Journal of Economic Behavior & Organization*, 146, 16–30. [96]

Appendix 5.A Robustness check

The punishment pattern that we report in the main text reveals that starting at 11-12 years of age, the majority of children enforce the egalitarian norm. In this robustness check, we further check whether the reported pattern is driven by different sample compositions of respective cohorts. For this purpose, we present multivariate regressions controlling for gender, cognitive abilities, parental income and pocket money. Table 5.3 shows four regression models. The dependent variable indicates whether a respective subject is a norm enforcer or not. In the first two models, the dependent variable "norm enforcer" is 1 if the subject does not punish on the egalitarian transfer but uses a total of at least 2 tokens for punishment on transfers smaller than 5 (egalitarian transfer), and 0 otherwise. The dependent variable for the latter two models is 1 if the subject does not punish on the egalitarian transfer but uses a total of at least 4 tokens on transfers smaller than 5, and 0 otherwise. The latter two models are included to provide a more robust definition of norm enforcers. The regressions show that in comparison to the 9-10 age cohort, being in an older age cohort significantly increases the likelihood of being a norm enforcer. This holds true if we control for gender, cognitive abilities, parental income and pocket money, thus confirming that there is an age-dependent increase of norm enforcement starting with 11-12 years of age.

Appendix 5.B Experimental instructions

The following section contains experimental instructions translated from Croatian.

5.B.1 Dictator

Let's start with the study. This study is conducted in a way that you play a game. Before the beginning of the game, I will explain the rules to you, so please listen carefully and do not talk. In the game, you will earn money, and all of the money that you earn will be given to you by us in an envelope. Each of you will receive (*value of 2 tokens expressed in kunas given the age cohort*), and additionally you will receive the amount of money that you earn in the game. There are no wrong answers in the game, but the amount of money that you earn will depend on your answers and the answers of other participants, so be very careful. Also, please do not look at your neighbor's sheet, but make decisions in the game by yourself.

The game is played in groups of 3 players. Each of you is the first player in your group. The second and third player from each group are also students of your age. I will not tell you who they are, neither will I tell them your names. This will remain a secret. The game is played with tokens. Each token is worth (*value of one token expressed in kunas given the age cohort*). This means that for each token that you have at the end of the game, you will receive those (*value of one token expressed in kunas given the age cohort*). This holds for you, and also for other players in the game.

In the game, as first players each of you have 10 tokens. Your task in the game is to divide those 10 tokens between you and the second player. You can divide the 10 tokens between yourself and the second player in any way that you want. The second player cannot influence your decision, and at the end of the game the second player will only have those tokens that you gave him.

This is how your decision sheet looks. Here you will decide how to divide the tokens (*show first player's decision sheet on the poster and give 3 transfer examples*, see Fig 5.7). But before you make your decision about the division of tokens between yourself and the second player, you have to know what the role of the third player is. The third player, if he chooses, can affect the number of tokens that you will have at the end of the game. Let's see how (*following on the poster while explaining verbally*, see Figure 5.8). At the beginning of the game, you have to divide tokens between yourself and the second player. Let's imagine that you leave 8 tokens for yourself, and you give 2 tokens to the second player. The second player will keep those 2 tokens at the end of the game. Whether you

will keep all of your tokens depends on the third player. The third player has 5 tokens. From those 5 tokens, if he chooses, he can use 2 tokens as deduction tokens. With those 2 tokens, he can reduce the number of your tokens. For each deduction token that he uses, your number of tokens will be reduced by 3. Let's see this on the poster. If the third player does not use any of his tokens as a deduction token, he is left with all of his tokens, and you are left with all of your 8 tokens that you left for yourself. If the third player uses 1 deduction token, he is left with 4 tokens, and the number of your tokens is reduced by 3. You left 8 tokens to yourself, and now you will have 8 minus 3, so you will have 5 tokens. If the third player uses 2 deduction tokens, he is left without those 2 tokens, so he will have 3 tokens, and the number of your tokens is reduced by 6. You left 8 tokens to yourself, and now you will have 6 less, so you will have 2 tokens. Let's see more examples. (*Show all possible third player's choices on 2 examples of first player's transfers by using posters with calculations of the outcome. Explain that the first player cannot have less than 0 tokens at the end of the game*.)

Does anybody have any questions?

Ok, now before we start the game, you have to answer three questions. These questions will help you to better understand the game, and it will show us if you have understood the game. If something is not clear, please ask. We will explain again if you have problems with correctly answering the questions. Now please turn the first page of your sheet where it says control questions and start answering. When you are done, lay down your pencil and wait, do not turn the pages any further. (*Wait until everybody in the classroom has answered the questions*.)

Now turn to the page of the game, and divide the tokens as you wish.

5.B.2 Punisher

Let's start with the study. This study is conducted in a way that you play a game. Before the beginning of the game, I will explain the rules to you, so please listen carefully and do not talk. In the game you will earn money, and all of the money that you earn will be given to you by us in an envelope. Each of you will receive (*value of 2 tokens expressed in kunas given the age cohort*), and additionally you will receive the amount of money that you earn in the game. There are no wrong answers in the game, but the amount of money that you earn will depend on your answers and those of other participants, so be very careful. Also, please do not look at your neighbor's sheet, but make decisions in the game by yourself.

The game is played in groups of 3 players. Each of you is the third player in your group. The first and second player from each group are also students of

your age. I will not tell you who they are, neither will I tell them your names. This will remain a secret. The game is played with tokens. Each token is worth (*value of one token expressed in kunas given the age cohort*). That means that for each token that you have at the end of the game, you will receive those (*value of one token expressed in kunas given the age cohort*). This holds for you, and also for other players in the game.

I have already played the game with the first and second players. Each first player in the group received 10 tokens. The first player had to divide those tokens between himself and the second player. He could divide those tokens as he wished, and the second player had no influence on his decision.

Let's look at one example (*following on the poster while explaining verbally*, see Figure 5.8). So, these are the 10 tokens that the first player had to divide. This is one example of how the first player could have divided those tokens between himself and the second player. Here, the first player left 8 tokens for himself, while he gave the second player 2 tokens. The second player will keep those 2 tokens at the end of the game. Whether the first player will keep all of his 8 tokens depends on you. Here we can see how. Each of you is the third player in your group, and each of you has 5 tokens. From those 5 tokens, if you wish, you can use 2 tokens as deduction tokens. With those 2 tokens, you can reduce the number of tokens of the first player. For each deduction token that you use, the number of the first player's tokens decreases by 3. And he is aware of this fact. If you do not want to use any of the tokens, you will be left with all 5 tokens, and the first player will also be left with all of his tokens. If you use 1 of your tokens as a deduction token, you will be left with 4 tokens, and the amount of first player's tokens will decrease by 3. How many tokens will he be left with? He had 8 tokens, now he has 3 less, so he will be left with 5 tokens. If you use 2 of your tokens as deduction tokens, you are left with 3 tokens, and the amount of the first player's tokens will decrease by 6. How many will he be left with? At the beginning he had 8, now he has 6 less, so he will be left with 2 tokens. Let's now look at an example of how exactly the game looks. (*Show all possible third player's choices on 2 examples of first player's transfers by using the third player's decision sheet as a poster*, see Figure 5.5. *Explain that the first player cannot have less than 0 tokens.*)

Does anybody have any questions?

Ok, now before we start the game, you have to answer three questions. These questions will help you to better understand the game, and it will show us if you have understood the game. If something is not clear, please ask. We will explain again if you have problems with correctly answering the questions. Now

please turn the first page of your sheet where it says control questions and start answering. When you are done, lay down your pencil and wait, do not turn the pages any further. (*Wait until everybody in the classroom answered the questions*.)

Now we will start with the game. As I have said, the first player gave to the second player as many tokens as he wanted, from 0 to 10. But you still do not know how many he gave him, and I will not tell you yet. When you open the decision sheets, you will see all of the possibilities of how the first player could have divided the tokens, with one possibility on each page. I want you to state how many deduction tokens you would use in each situation. So, first you will see a situation where the first player gave 0 tokens. You decide how many deduction tokens you would use in that situation. Then you turn to the next page, where you will see a situation where the first player gave 1 token. You decide how many tokens you would use in that situation. Then you turn to the next page, and so on. After you decide how many deduction tokens you would use on all possible divisions of the tokens, we will check how the first player actually divided the tokens. We will see how many deduction tokens you decided to use on that division, and then apply only that decision. (*Short pause.*) Ok? So, you decide for every possible first player's division how many deduction tokens you would use. But only one of your decisions will determine the payoff at the end of the game. You are still not aware which decision it is, so carefully decide on each possible division. Are there any questions? (*Explain the procedure again in case somebody does not fully understand.*) Now open the first page, and start the game.

Appendix 5.C Additional tables and figures

Table 5.3. Probit regressions estimates of age cohort effects

Variables	Dependent variable: norm enf. ($\geq$ 2 tokens)		Dependent variable: norm enf. ($\geq$ 4 tokens)	
	(1)	(2)	(3)	(4)
Basis: 9-10 age cohort				
11-12 age cohort	0.239*	0.233*	0.252**	0.243**
	(0.128)	(0.130)	(0.121)	(0.118)
13-14 age cohort	0.299**	0.305**	0.329***	0.273**
	(0.119)	(0.130)	(0.113)	(0.118)
15-16 age cohort	0.257**	0.347**	0.274***	0.340**
	(0.113)	(0.150)	(0.106)	(0.134)
17-18 age cohort	0.398***	0.473***	0.379***	0.481***
	(0.102)	(0.134)	(0.097)	(0.119)
Gender (1 if male)		-0.027		-0.086
		(0.082)		(0.081)
Cognitive abilities		0.030		0.016
		(0.041)		(0.040)
Pocket money (1 if yes)		-0.054		-0.030
		(0.104)		(0.104)
Family income		0.029		0.027
		(0.028)		(0.028)
Observations	180	163	180	163

Notes: The table presents probit regressions using dummy variable for being a norm enforcer (= 1 if subjects is a norm enforcer) as the dependent variable. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The reported coefficients represent average marginal effects evaluated at sample means. Norm enforcer (min. 2 tokens) is 1 if a subject does not punish on the transfer of 5 (egal. transfer) but uses a total of at least 2 tokens for punishing transfers smaller than 5, and 0 otherwise. Norm enforcer (min. 4 tokens) is 1 if a subject does not punish on the transfer of 5 but uses a total of at least 4 tokens for punishing transfers smaller than 5, and 0 otherwise. The value for cognitive abilities was normed for each of the age cohorts using the dictators', recipients' and punishers' test results to avoid confounding age and cognitive abilities effects. Family income is measured in 7 categories; it takes a value of 1 if family income < 2000 HRK, 2 if family income $\geq$ 2000 and < 4000 HRK, 3 if family income $\geq$ 4000 and < 6000 HRK,..., 7 if family income $\geq$ 12000 HRK. Models (2) and (4) include controls for high school program fixed effects.

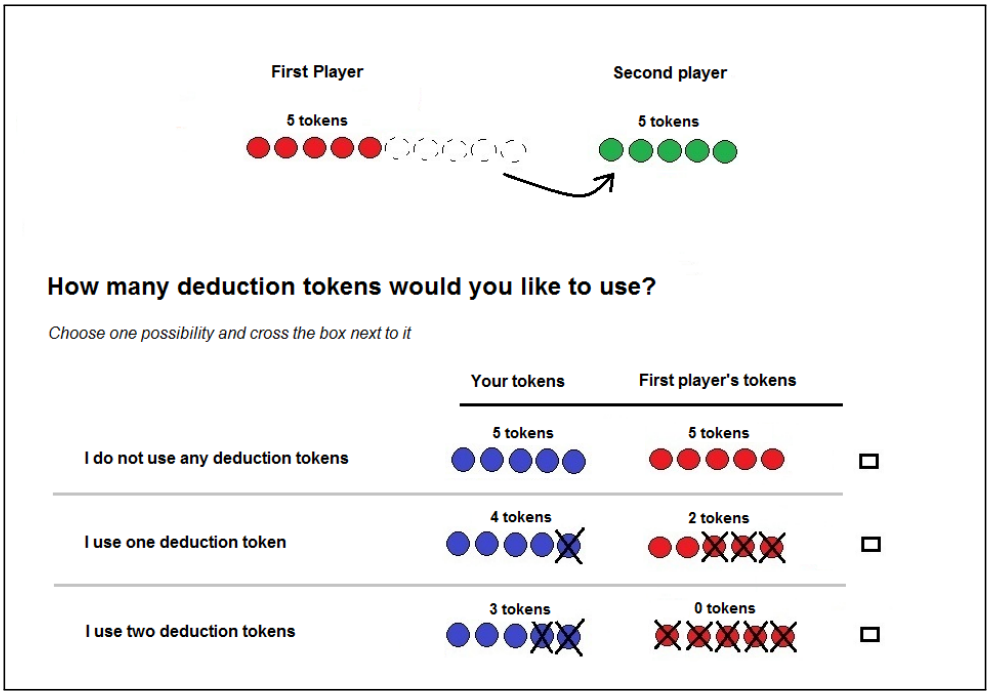

Figure 5.5. Punisher's decision sheet: example of dictator's transfer of 5 tokens

Notes: Translated from Croatian.

Figure 5.6. Experimental session

Divide the tokens between yourself and the second player

Cross the box next to the division you chose

	Your tokens	Second player's tokens
☐	10	0
☐	9	1
☐	8	2
☐	7	3
☐	6	4
☐	5	5
☐	4	6
☐	3	7
☐	2	8
☐	1	9
☐	0	10

Figure 5.7. Dictator's decision sheet

Notes: Translated from Croatian.

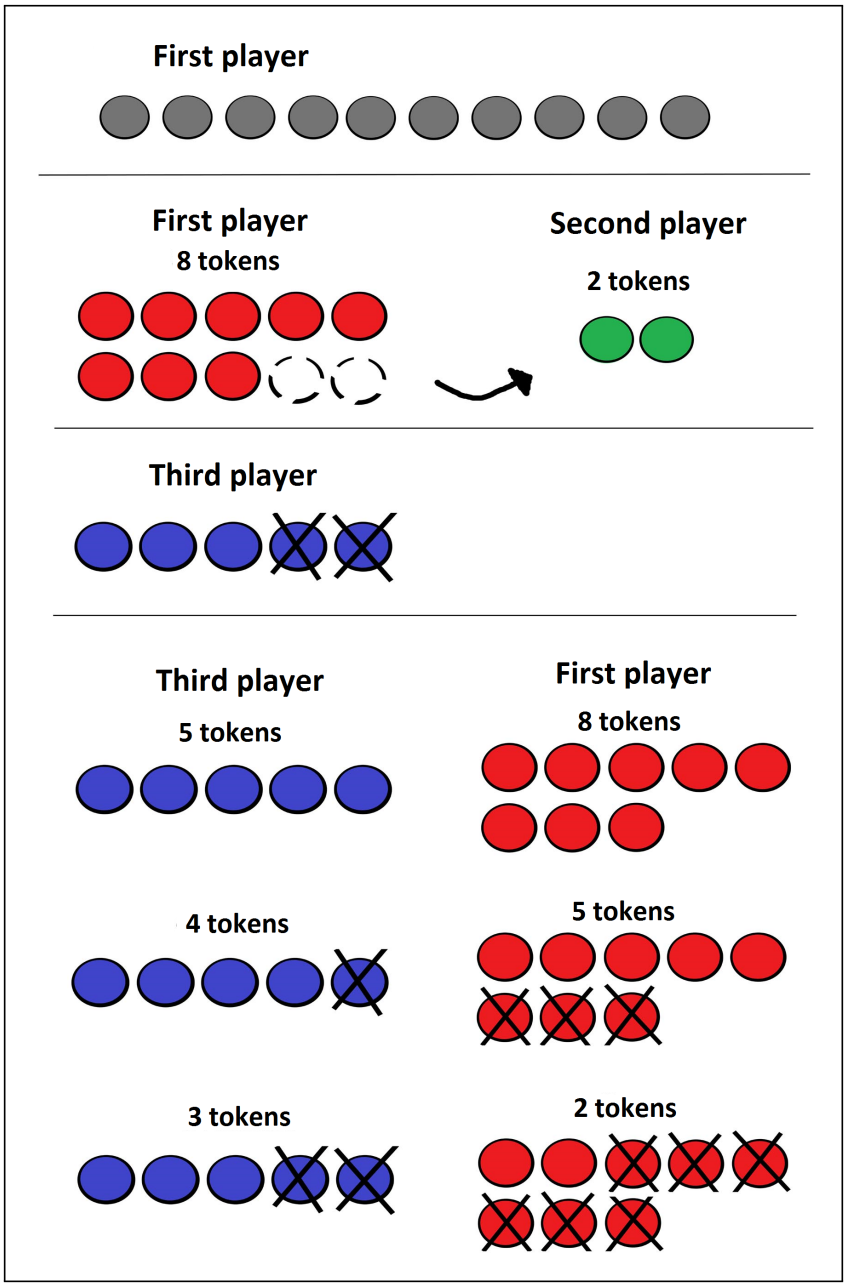

Figure 5.8. Poster used during oral instructions for the main explanation of the game

Notes: Translated from Croatian.