Essays on Moral and Ethical Behavior in Experimental Economics

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2020

vorgelegt von

Eugenio Verrina

aus

Genua (Italien)

Referent: Prof. Dr. Bettina Rockenbach

Koreferent: Prof. Dr. Dr. h.c. Christoph Engel

Tag der Promotion: 14.07.2020

Acknowledgements

I would like to start by thanking the MPI for Research on Collective Goods and the University of Cologne for giving me the chance to pursue my doctoral studies and for providing the financial support to my research activities. The decision to embark in a PhD came almost naturally after my studies in Trento and I owe this mostly to Matteo Ploner who was a great mentor and has since become a friend. I feel very fortunate to be able to accomplish this goal and I am proud to share this achievement with everyone who has helped me along the way. My supervisors Bettina Rockenbach and Christoph Engel offered me great support and guidance throughout my PhD. I am extremely grateful to the IMPRS-Uncertainty School and my colleagues from Berlin and Jena for everything I learned during our Summer Schools and Workshops and for putting together such a beautiful mix of researchers. I want to thank all my colleagues at the University of Cologne for their exceptional support and for offering a very stimulating environment, a special thank goes to Thomas Lauer and Lukas Wenner for their advice and assistance. The MPI has become like a second home to me, not only because I spent there many late nights and weekends, but most of all because of the people I have met there. The "Profis" (Cornelius Schneider, Maj-Britt Sterba and Martin Sternberg), surely contributed a lot to make me feel at home. I am also extremely grateful to have worked with my co-authors and "big brothers" Zvonimir Bašić and Adrian Hillenbrand. All the colleagues at the MPI would deserve a separate mention, but I want to express my gratitude especially to Amalia Álvarez Benjumea, Stefania Bortolotti, Philip Brookins, Claudia Cerrone, Lars Freund, Leonard Hoeft, Jerome Olsen, Angelo Romano, Ali Seyhun Saral and Fabian Winter who offered me their tips and support when I needed them. I am also very grateful to Matthias Sutter for always taking the time to give me his prompt and very helpful advice. An additional special thank goes to the all the personnel in the administration, IT, library and scientific services at the MPI for making research as smooth and easy as possible. During my PhD I had the wonderful chance to spend a visiting period at the University of Zurich which contributed a lot to enlarge and refine my research agenda. I am very thankful to Roberto Weber and the great researchers that I have met there.

It is hard to imagine how I would have been able to navigate these years without the help of my friends and family. I have learned that the best friendships are those that withstand time and distance. My first and longest friends are certainly my brothers, together we developed our personalities and characters, different but always united. My parents always enabled me to go after my ambitions and have prepared me to meet all the challenges I faced long before I could realize it. My mom injected me with great curiosity and open-mindedness, my dad taught me dedication and commitment and both showed me with their own example the value of hard work. Finally, these years I have been accompanied by Nives who gives sense to much of what I do and always overwhelms me with her love.

Contents

List of Figures

List of Tables

Introduction

Despite the common assumption of selfish, rational utility-maximizing agents in standard economic models, a surprising amount of real-world economic interactions relies or is strongly influenced by some form of non-selfish, i.e., moral, ethical or normative, behavior. This is not really surprising. Real people take into account others and follow norms when they take decisions. We pay taxes to finance public goods despite the far too low inspection probability. We donate money to causes that will not benefit us directly. On a more aggregate level, we hold organizations accountable for their ethical record. In markets, reciprocity between employers and employees can keep wages above equilibrium levels, but can also prevent market breakdowns. These are just some instances of how non-selfish behavior can shape economic outcomes.

Of course, economists did not completely ignore this and, today, considerable evidence in favor of the importance of non-selfish behavior has been gathered. With it our understanding of these phenomena has become increasingly refined and has been distilled into formal models that look at such behavior from various different angles. Some of the first theories of non-selfish behavior were "outcome-based", i.e., they incorporated the distribution of monetary payoffs in the utility function of an individual (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). These models of "social" preferences have gone a long way to explain a wide variety of findings, but were necessarily rather naive on some aspects of human motivation. A second strand of theories that were developed almost concurrently acknowledged the role of intentions and emotions in economic behavior using the tools of psychological game theory (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009). They introduced the notion of reciprocity in economics (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) and emphasized the role of guilt in non-selfish behavior (Battigalli and Dufwenberg, 2007). A third group of theories started to branch out slightly later with the aim to incorporate concerns for self and social-image (Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009). In these models, individuals want to signal to others and to themselves that they are moral or virtuous.

At least since the seminal paper by Dana et al. (2007), however, a quickly expanding body of evidence has directed the economic literature on non-selfish behavior in yet a new direction. Dana et al. (2007) show how non-selfish behavior can crumble when apparently negligible aspects of the decision are changed. While this does not disqualify the theories mentioned above, many other studies have documented that people exploit excuses and justifications to behave selfishly

(see, e.g., Konow, 2000; Haisley and Weber, 2010; Exley, 2015; Grossman and Van Der Weele, 2017). This evidence indicates that people are *motivated* thinkers who want to feel or appear moral (Bénabou and Tirole, 2016; Gino et al., 2016). The intuition that people manipulate their beliefs about the world is supported by psychological theories of motivated reasoning (Shafir et al., 1997; Ditto et al., 2009) and can also be linked to the need to reduce cognitive dissonance (Festinger, 1962), which has found its way in other fields of economics (Akerlof and Dickens, 1982). Still, there is yet no coherent framework that has established itself to account for this behavior.

Chapter 1 of my dissertation (based on joint work with Adrian Hillenbrand), contributes to the understanding of motivated reasoning in prosocial behavior by studying the role of *narratives* as one of its key ingredients. Psychological theories define narratives as tools people use to make sense of the world around them and of their own behavior (Bruner, 1991; McAdams, 1988). In economics, narratives have been used to explain fluctuations in markets (Shiller, 2017) and also broader historical phenomena such as the rise and fall of the Soviet Union (Akerlof and Snower, 2016). In the context of prosocial behavior, Bénabou et al. (2018) model them as stories or justifications that help maintaining a positive self and social-image.

We study how positive and negative narratives influence behavior in a modified dictator game where an individual has to distribute a given amount of money betweenherself and a recipient. We conceptualize narratives as arguments targeting the perception of the appropriateness of an action and the deservingness of a recipient. Positive narratives are arguments in favor of the prosocial action, i.e., sharing the pie equally with the recipient. Negative narratives, on the other hand, are arguments justifying the selfish action, i.e., giving nothing to the recipient. We use arguments that subjects themselves provide in previous experimental sessions in justification of their own choice to construct two treatments where subjects are confronted with either positive or negative narratives and compare this to a baseline treatment with no narratives. We isolate the causal effect of narratives as providing reasons for either the selfish or the prosocial action by showing subjects in all treatment conditions a distribution of choices made in similar dictator game experiments. A key aspect of our design is that we control for the prosocial inclination of individuals by measuring their Social Value Orientation, as heterogeneity in this dimension plays an essential role in theories of prosocial behavior (see, e.g., Bénabou and Tirole, 2006) and recent empirical evidence confirms that individuals' prosocial preferences greatly vary (Falk et al., 2018). Then, we develop a theoretical framework and spell out predictions about how externally supplied positive or negative narratives affect different prosocial types.

Our main findings are that, in line with our predictions, positive narratives increase giving of selfish types substantially, compared to a baseline with no narratives. However, contrary to our predictions, negative narratives have a differential effect. Prosocial types decrease their giving, while selfish types give more than in the baseline. We argue that narratives offer a benchmark for social comparison, besides influencing perceptions of deservingness and appropriateness. Subjects are influenced by narratives and, at the same time, compare themselves with the narrator. They

seek to match the behavior of a prosocial narrator and want to distinguish themselves from a selfish narrator. This explanation can account for the complete pattern of results, including the differential effect, and is supported by additional results on the extensive and intensive margin of giving. The findings from Chapter 1 provide some of the first empirical insights on how narratives in the domain of prosocial behavior work. We find evidence suggesting that they do, indeed, target motivated reasoning thorough perceptions of deservingness and appropriateness, but may also evoke a vivid comparison with the narrator.

Chapter 2 of my dissertation deals with another essential mechanism of motivated reasoning: *self-serving beliefs*. As mentioned above, a large body of literature has researched the conditions under which people display moral behavior and those under which they act in their own self-interest. One of the ways people resolve the tension between feeling moral and acting in their own interest is by distorting their beliefs. In a recent paper, Di Tella et al. (2015) investigate the formation of self-serving beliefs justifying unfair behavior in a "corruption game". In Chapter 2, I replicate their study with few changes in the design, but fail to reproduce their findings.

In the "basic" version of the game I study, an Allocator decides how to distribute an endowment between herself and a Seller. The Seller decides how to convert this endowment into monetary payoffs. She can choose between a high conversion rate or a low conversion rate. If she chooses the latter, she receives a bribe which is independent from how the Allocator distributes the endowment. The Allocator is asked whether she thinks the Seller has taken the bribe and how many Sellers in her sessions did so. The authors can identify self-serving beliefs by comparing a treatment in which the ability of Allocators to redistribute the endowment is strongly restricted with one in which Allocators can redistribute a larger share of the endowment. According to their hypothesis, Allocators who can redistribute more will think that more Sellers take the bribe to justify redistributing a larger share of the endowment to themselves. While the authors find this to be the case, my results point, if anything, into the opposite direction. In fact, Allocators in my study have much more negative beliefs about Sellers. If taken together with a desire to maintain a positive self-image, this can potentially explain why I fail to replicate the original results. This chapter uncovers the sensitivity of self-serving beliefs and exposes some of the challenges for the formal modeling of these constructs.

Chapter 3 (which is based on joint work with Zvonimir Bašić) investigates the relationship between *personal norms*, social norms and image concerns. This chapter takes an alternative, parallel path compared to the theories described above to model non-selfish behavior. Social norms have deep roots in the economics literature (Akerlof, 1976) as well as in that of psychology and sociology (Schwartz, 1977; Cialdini et al., 1991; Bicchieri, 2005). They can be braodly conceptualized as societal prescriptions about how one ought to behave. While social norms have been used by economists to explain non-standard behavior in a wide variety of settings (Akerlof, 1980; Lindbeck, 1997; Fehr et al., 1998; Lindbeck et al., 1999; Fehr and Fischbacher, 2004; Gächter and Schulz, 2016), personal norms almost never feature in the literature although they are a well-established construct in other social sciences (see, e.g., Schwartz, 1973, 1977;

Cialdini et al., 1991; Bicchieri, 2005).

Our study aims at putting personal norms under the spotlight and highlighting their relevance for economic behavior alongside social norms. As a first step, we propose a simple utility framework in which people care about their monetary payoff, social norms and personal norms. We also posit that the weights put on the two norms can vary depending on the context in which a decision is taken. We then design a novel two-part experiment which allows us to investigate the predictive value of personal norms as well as social norms across four economic games. In the first part of the experiment, we elicit subjects' social and personal norms with a symmetric procedure via an online survey for four economics games. In the second part of our experiment, which takes places approximately four weeks later in the lab, subjects play the four games we elicited the norms from. We assess the influence of situational factors with a treatment in which choices are publicly observable. Our hypothesis is that this will increase subjects' social image concerns boosting the relation between social norms and behavior.

We start by clarifying that, while personal and social norms are related, there is substantial heterogeneity at the individual level between the two. We proceed to show that personal norms are a strong predictor of behavior across all games in our baseline treatment where choices are private. This also holds when social image concerns are high, which indeed increases the predictive value of social norms. A model comparison exercise confirms that adding personal norms significantly increases the predictive fit of a model which considers only social norms and monetary payoffs. This further strengthens our claim that personal norms are complementary to social norms in predicting behavior. After a thorough robustness check of our findings, we reproduce our results on the relationship between the two norms with a different sample and show that the two constructs are systematically different also for several other games and in real-life economic situations. Taken together, this evidence corroborates the role of personal norms as indispensable drivers of non-selfish behavior. Indeed, ignoring them can lead to misguided interpretations of behavior and ill-designed policy interventions.

While the first three chapters of my dissertation answer more fundamental questions about non-selfish behavior, in Chapter 4, I tackle an applied ethical problem. Decision-makers in firms and organizations often face the trade-off between higher profits and potential negative externalities. Cutting the costs on activities that decrease health or environmental risks will increase profits, but exposes others to negative externalities. One of the key features of organizations is that decisions are taken in teams in order to aggregate different skills and sources of knowledge. I study whether the informational asymmetry about negative externalities originating in team decision-making can lead to more unethical behavior. I investigate whether this comes about because "experts", i.e., the better informed decision-makers, exploit their informational advantage to act more unethically, or because they ignore their private information and quietly agree to implement actions they know could be harmful.

I run an experiment in which two subjects have to decide between two options of which one is more profitable for them, but entails a risk for another passive subject. They receive signals

about the actual risk of this negative externality and one of them receives an additional, more precise signal (the expert) compared to the other (the non-expert). The design is conceived to isolate three behavioral channels that arise due to asymmetric information and could lead to unethical behavior. My results have two sides. On the bright side, I find that experts do not behave more unethically when the decision is delegated to them. On the contrary, experts seem to condition their decision on the information they share with the non-expert and do not initiate more unethical behavior. Additional data suggest that this is due to an enhanced feeling of responsibility. However, on the dark side, they do not intervene to avoid unethical outcomes, thereby ignoring their private information. This hints at an omission-commission asymmetry (Spranca et al., 1991) that allows experts to find an excuse for their unethical behavior (Mazar et al., 2008). Overall, high negative externalities are generated despite the presence of the expert. This study is an example of the relevance of motivated reasoning in applied settings. My results suggest that organizations should prevent experts from adopting a passive role in the decision process, if they want to curtail unethical behavior.

In a nutshell, the work presented in this dissertation contributes to the fundamental discourse in the realm of motivated moral and norm-guided behavior and also answers applied ethical questions. It investigates when people behave morally, what guides their actions and how they foul themselves into believing or making others believe to be moral while acting selfishly.

Chapter 1

The Differential effect of Narratives on Prosocial Behavior

Imagine that for some days you have seen a beggar on your way to work. As you pass by today, you reach into your pocket to get some change. While doing so, you remember what a colleague told you the day before. He stated that most of these people are not really needy, but have simply chosen to live soaking up money from people who work hard. Besides, according to your colleague, the beggar will spend all the money you give him on alcohol and drugs; he deserves no consideration at all. Now imagine your colleague telling you instead that rising inequality is destroying our society and that the government does not do enough for people in need. He says we should all fight against the unfairness of this wicked capitalistic system. Will you give something to the beggar after recalling one of the two stories? Will you give him more or less than what you had picked from your pocket in the beginning? Will you react differently based on your first tendency to give or not to give something?

Theoretical accounts of motivated moral reasoning (Ditto et al., 2009) emphasize people's deep need to justify their moral behavior not only to others, but especially to themselves. From a fully rational standpoint, these justifications could reflect pieces of evidence an individual uses to inform her choice. However, cognitive dissonance theory (Festinger, 1962) indicates how such reasons can often be used beyond that to resolve tensions between beliefs and actions (Akerlof and Dickens, 1982).[1] In our opening illustration, the tension between a self-interested and a prosocial option can be resolved differently, depending on the story one is told or recalls. We will call these stories that come in the form of rationales or justifications *narratives*. The notion of narratives is deeply grounded in psychological theories (Bruner, 1991; McAdams, 1988), where they serve as tools people use to construct their own account of the world. As such, narratives accompany nearly all our decisions, often playing a decisive role in shaping them. Their relevance for economic outcomes has recently received growing attention. Narratives help explain fluctuations

[1] Epley and Gilovich (2016) make a very similar point in their discussion of the mechanics behind motivated reasoning in general.

in markets (Shiller, 2017) and also broader historical phenomena (Akerlof and Snower, 2016). Recent theoretical work by Bénabou et al. (2018) has contributed to the understanding of how narratives affect moral or prosocial behavior.[2] The authors develop a model where individuals with self and social image concerns produce and consume narratives as signals complementing their actions.[3] Unfortunately, naturally occurring data do not allow to isolate the effect of these moral arguments, since they are often bundled together with other types of information. This poses serious challenges in getting at the causal effect of narratives as rationales in favor of a certain behavior.

In this chapter, we test how providing narratives affects prosocial[4] behavior by leveraging the control of a laboratory experiment. In particular, we look at the impact of positive and negative narratives, as defined by Bénabou et al. (2018). Positive narratives are arguments endorsing moral or prosocial behavior, e.g, by highlighting the presence of a norm or potential reasons supporting it. Negative narratives, on the other hand, are arguments justifying immoral or selfish behavior and can operate through various mechanisms; they can, e.g., downplay the negative externalities of an action or alter the normative expectations pending on the decision-maker. By controlling for the prosocial inclination of individuals, we analyze whether positive or negative narratives affect different types of individuals differently. Heterogeneity in this dimension plays an essential role in theories of prosocial behavior (see, e.g., Bénabou and Tirole, 2006) and recent empirical evidence confirms that individuals' prosocial preferences greatly vary (Falk et al., 2018).

In our experiment, subjects play a dictator game in which they decide how to share a given amount of money with another anonymous participant. In our two treatment conditions, they are shown either negative or positive narratives while making their choice. Narratives in the NEGATIVE condition are arguments in favor of the selfish action, i.e., giving nothing to the other participant, while narratives in the POSITIVE condition are arguments in favor of the prosocial action, i.e., splitting the money equally.[5] We capitalize on arguments subjects used in previous experimental sessions for justifying their own choice to construct our treatments. This confers greater internal validity to our experimental design and allows us to systematically study the

[2]Bénabou et al. (2018) also discuss "imperatives", i.e., statements issued by a moral authority dictating to follow a given behavior, as an alternative way to convey moral arguments. The authors present a model, in which a principal who cares about the welfare of an agent can choose to send her either a narrative or an imperative. We focus on settings in which no such authority exists or in which she does not have enough persuasive power to issue an imperative.

[3]Foerster and van der Weele (2018a) work out a similar model where two agents with social image concerns can exchange signals about the social returns to an investment in a public good in a simultaneous pre-play communication phase. Their model generates a set of predictions about the use of the signals which are comparable with Bénabou et al. (2018) for what concerns the focus of this chapter. In a companion paper, Foerster and van der Weele (2018b) also test their model.

[4]We focus on prosocial behavior as an important component of moral behavior. As opposed to prosocial behavior, we equate immoral behavior to selfish behavior.

[5]Krupka and Weber (2013) provide compelling empirical evidence that the equal split is indeed considered to be the most socially appropriate behavior in the dictator game. In this sense, what we label as the prosocial action would correspond to the social norm, while what we call the selfish action would be the strongest possible deviation from the social norm or the most inappropriate behavior. As hinted in our behavioral predictions (see Section 1.2.2), our hypotheses also hold in a social norms framework.

effect of the content of narratives, i.e., their appeal to the selfish or the prosocial action. We compare our two treatments to a BASELINE condition with no narratives. Importantly, we keep empirical expectations across all our conditions constant by showing subjects a distribution of choices made in similar dictator game experiments. This ensures that our treatment manipulations do not carry any valuable empirical information about the relative frequency of choices. We thus isolate the causal effect of narratives as providing or highlighting reasons for either the selfish or the prosocial action.

A key feature of our design is that it allows us to explore how heterogeneous prosocial concerns interact with positive and negative narratives by using subjects' Social Value Orientation (SVO). We look at how individuals who are more or less prosocial react to the narratives we present them. To that end, we provide a theoretical framework to illustrate how externally supplied narratives influence giving of types with different prosocial orientations and derive simple hypotheses to benchmark our experimental results. Narratives, in our setting, are arguments targeting the perception of recipients' deservingness as well as the appropriateness of giving or not. According to our predictions, positive narratives should increase aggregate giving, while negative narratives should decrease it. The effect should go in the same direction for all social[6] types and should be stronger for prosocial types who see a negative narrative and selfish types who see a positive narrative.

Our main results are that positive narratives increase giving, while negative narratives have a *differential effect* on different social types. In line with our predictions, types across the whole spectrum increase their giving in the POSITIVE condition, with selfish types displaying the largest reaction. However, in the NEGATIVE condition, prosocial types decrease their giving, while selfish types increase their giving. This result is at odds with our hypotheses, according to which the same narrative cannot cause certain types to increase and other types to decrease giving.

We offer two potential explanations for this effect. According to the first, narratives - both positive and negative - enhance the salience of the moral decision, thus making it harder for subjects to behave selfishly. According to the second explanation, narratives provide a benchmark for social comparison. Subjects are, thus, induced to compare themselves with the narrator. They seek to match the behavior of a prosocial narrator and want to distinguish themselves from a selfish narrator. Our social comparison explanation can account for the complete pattern of results, including the differential effect, and is supported by additional results on the extensive and intensive margin of giving. This suggests that narratives may evoke a vivid comparison with the narrator beyond targeting perceptions of deservingness and appropriateness. We believe that capturing this motive can lead to important insights in prosocial behavior. From a practical standpoint, our results suggest that organizations and institutions can promote prosocial outcomes by confronting people with different narratives, positive or negative, depending on their predisposition. The evidence we present indicates that narratives have the potential to increase prosocial behavior especially among those who would be less inclined to behave prosocially ex

[6] We use the term "social" types to indicate all individuals with different prosocial orientations and the terms "prosocial" (or prosocials) and "selfish" to refer to individuals with high or low prosocial concerns.

ante.

1.1 Related literature

Our work resonates with the growing interest in the role played by narratives (Bénabou et al., 2018; Foerster and van der Weele, 2018a; Shiller, 2017; Akerlof and Snower, 2016) and, more generally, in the role motivated reasoning plays in shaping economic interactions (Karlsson et al., 2004; Epley and Gilovich, 2016; Bénabou and Tirole, 2016; Golman et al., 2016; Gino et al., 2016; Carlson et al., 2018; Saucet and Villeval, 2018). Our work is also closely linked to experimental studies on phenomena of so-called moral wiggle room (Dana et al., 2007; Larson and Capra, 2009; Matthey and Regner, 2011; van der Weele et al., 2014; Feiler, 2014) and to the wider literature investigating self-serving judgments of fairness or morality (Konow, 2000; Hamman et al., 2010; Shalvi et al., 2011a; Wiltermuth, 2011; Rodriguez-Lara and Moreno-Garrido, 2012; Bicchieri and Mercier, 2013; Gino et al., 2013; Shalvi et al., 2015; Exley, 2015) and self-serving beliefs (Haisley and Weber, 2010; Chance et al., 2011). The main result one can draw from this huge body of evidence is that prosocial behavior is highly sensitive to the specific context in which choices take place, and that people often tweak the evidence in their favor in conscious and unconscious ways. Our work contributes to this growing literature by providing evidence on how people react to externally provided narratives and by analyzing how heterogeneity in prosocial concerns affects behavior in this context.

Andreoni and Rao (2011) study a setting in which Receivers and Dictators in a dictator game can communicate with each other. They find that giving increases whenever Receivers can say something. Whereas, if only Dictators have the word, giving decreases. We investigate a setting in which Dictators are exposed to arguments coming from other Dictators, who behaved either prosocially or selfishly. People are constantly exposed to such arguments both in their professional and private life. We systematically study their effect on prosocial behavior. Similarly, Mohlin and Johannesson (2008) find a positive effect of one-way communication from the Receiver to the Dictator and also from past Receivers to Dictators. Differently from these and other studies of communication in economic games (see, e.g., Bohnet, 1999; Charness and Dufwenberg, 2006), we do not look at the effect of communication between parties involved in the game. Instead, we analyze the effect of justifications or rationales, i.e., narratives, that individuals provide for their own choice on the behavior of other individuals facing the same decision.

Other work has looked at how contextual factors, e.g., frames (Brañas-Garza, 2007; Dreber et al., 2013) or social information (Krupka and Weber, 2009; Gino et al., 2009; Cappelen et al., 2013, 2017), influence prosocial behavior. We hold these channels constant and explicitly provide reasons, or narratives, for a certain action. This links our work to studies investigating the effect of moral reminders or recommendations on behavior (see, e.g., Galbiati and Vertova (2008) on obligations and Croson and Marks (2001) on recommendations, both in the public-good game, or Mazar et al. (2008) in the context of lying; further work by Bott et al. (2017) uses moral appeals in letters to tax payers). Most closely related to our study is an experiment by Dal Bó and Dal Bó

(2014), who look at the effect of moral suasion in the form of arguments issued by an authority[7], i.e., the experimenter, in favor of the socially optimal contribution in a voluntary contribution game. In contrast to them, we look at a non-strategic setting where narratives can only affect preferences and cannot act as coordination devices. Moreover, our messages do not come directly from the experimenter, but are naturally occurring reasons subjects in previous sessions provide for their choices. These features of our experimental design allow us to test systematically the effect of the content of narratives, i.e., their appeal in favor of the selfish or prosocial action. Last but not least, measuring prosocial concerns allows us to look at heterogeneous effects on different social types and to test the effect of what we call negative narratives more thoroughly.[8].

To achieve this goal, we use the SVO slider measure by Murphy et al. (2011) to measure subjects' social types. The SVO measure has been widely used in both psychology and economics to assess heterogeneity in individual motives in social and moral dilemmas (see Balliet et al., 2009, for a meta-study on SVO and cooperation in social dilemmas), e.g. in the public-good game (see e.g. Offerman et al., 1996). Other studies find that individuals scoring differently on the SVO measure exhibit different behavior also in other realms, such as intergroup conflict (Weisel et al., 2016), in vaccine-related behavior (Böhm et al., 2016), and in pay what you want settings (Krämer et al., 2017). Grossman and Van Der Weele (2017) study a setting where people can remain ignorant about harmful consequences of their actions, and find that the SVO measure confirms the sorting predictions of their model. In line with previous studies, we are interested in how heterogeneous prosocial concerns interact with our treatment manipulations. We find this to be indeed an important dimension to look at, since different types display not only quantitatively, but also qualitatively different reactions.

1.2 Experimental Design

1.2.1 Setup

Our experimental design consists of two main building blocks (see Figure 1.1): an online pre-study and a laboratory experiment. The laboratory experiment is subdivided in a modified dictator game and a questionnaire containing various ex-post measures. The online pre-study was conducted one week before the experiment.[9] The laboratory experiment was implemented in a between-subjects design with a Baseline and two treatment conditions (Positive and Negative), which varied only in the content of the narratives subjects saw. Below, we discuss the individual parts of the study in detail. Instructions for the laboratory experiment can be found in Appendix A.1.1.

[7]The moral suasion treatments in Dal Bó and Dal Bó (2014) is very close to the notion of imperatives in Bénabou et al. (2018). In this sense, our study and the one by Dal Bó and Dal Bó (2014) can be understood as testing the effect of narratives and that of imperatives, respectively.

[8]Dal Bó and Dal Bó (2014) find that messages explaining the game-theoretical prediction of zero contribution have no effect on contributions. However, baseline contributions are already quite low when they introduce this manipulation and there is hardly any room for a further decrease to take place.

[9]Subjects received the link to the pre-study one week before the experiment and had three days to complete it. Subjects generated a code that was used to match their responses from the pre-study with those from the lab.

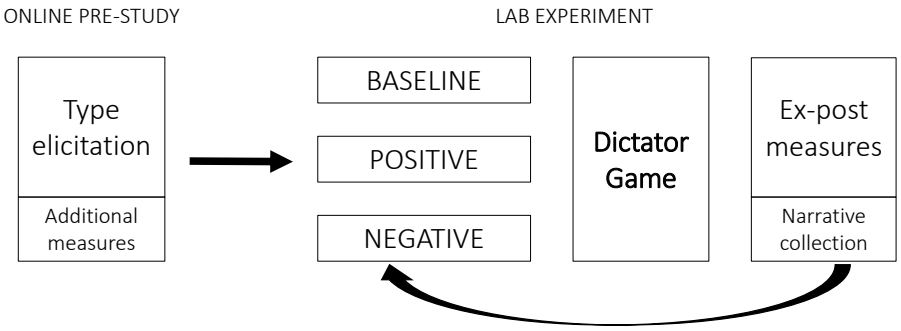


Figure 1.1: Experimental Design

Dictator game. The central part of our design is constituted by a simple dictator game (Kahneman et al., 1986). Dictators chose how to divide 10 € between themselves and an anonymous recipient (in intervals of 1 €). All subjects in the experiment decided under role uncertainty[10], i.e., each subject made her choice in the role of the dictator and roles were randomly assigned at the very end of the experiment.

Crucially, we fixed subjects' empirical expectations about the distribution of giving in the dictator game. This makes sure subjects could not take the narratives in our treatment conditions as signals about the empirical distribution of giving. Subjects in all experimental conditions were presented with a graph showing the distribution of dictator game giving in similar experiments (see Figure A1 in Appendix A.1.2). The graph displays data from Engel (2011) restricted to studies in which 10 units of currency were used. Subjects were told the graph displayed the distribution of choices other subjects had made in similar previous experiments.[11] The figure displays the typical bimodal distribution with modes at 5 € and 0 € and a sizeable mass in between. While holding empirical beliefs constant across our experimental conditions, the distribution does not clearly emphasize one allocation choice over the other.

Treatments. Participants were randomly allocated to one of three treatment conditions in a between-subjects design. In the Baseline condition, subjects only saw the distribution of dictator game giving described above. In the two treatment conditions, they were additionally

[10]Iriberri and Rey-Biel (2011) find that role uncertainty increases selfish choices. To the extent to which the increase is not excessive and does not interact with our treatment manipulations, this does not constitute a problem for our design.

[11]We used the following expression: "The figure shows the frequency of choices of participants in similar experiments in percentages."

shown two comments which subjects in the BASELINE condition had used to explain their choices. These are our narratives (see Appendix A.1.3). In the POSITIVE condition, subjects saw two comments in support of the equal split (giving 5 €), while in the NEGATIVE condition they saw two comments justifying selfish behavior (giving 0 €). Subjects were (truthfully) told that these were explanations other participants had given for their choices in similar previous experiments.[12] In the next paragraph, we explain how we collected and selected the narratives to devise our treatment conditions.

Narrative collection. After subjects had gone through all stages of the experiment, but before their final roles for the payment were revealed, they were given the opportunity, without any prior notice, to explain the reasoning behind their choice in the dictator game.[13] We used the explanations from the BASELINE condition to build the set of narratives subjects saw in the POSITIVE and NEGATIVE condition. Three independent raters, who were blind to the research question, evaluated the narratives along several dimensions. First, they were asked whether it was possible to understand what a subject had chosen in the dictator game from his or her comment and, if so, which was the most likely choice (0,1,2, etc.). Raters also evaluated how convincing they perceived the narrative to be (on a 7-point Likert scale).[14]

We then selected the most convincing narratives in support of giving 0 € and in support of giving 5 € (using average ratings). We excluded narratives which were particularly long or repetitive. We selected four positive and four negative narratives. Each individual in the two treatment conditions saw two randomly selected narratives (at individual level). We take these steps, on the one hand, to prevent our results from depending on a single item and, on the other, to increase the probability of subjects indeed being treated by at least one narrative. See Appendix A.1.3 for the list of selected narratives.

Type elicitation. As mentioned above, the online pre-study was conducted one week prior to the laboratory experiment to avoid contamination across the two. The purpose of our online pre-study was to measure subjects' prosocial concerns. Our main measure of a subject's social type is the SVO slider measure (Murphy et al., 2011). Subjects are confronted with 6 choices where they have to trade off their earnings with those of another subject under different budget constraints. From these choices, the so-called SVO angle is constructed, which represents the relative weight subjects put on the payoff of others compared to their own. Subjects with an SVO angle of 0° care only about their payoff, while those with an SVO angle of 45° weigh their payoff and that of the other subject equally. Types with an SVO angle below 25° are generally classified as selfish and those above as prosocials. Earnings in this task are determined by forming

[12] We used the following expression: "Here are two explanations (*Begründungen*, in German), which other participants gave for their choice."

[13] The exact wording was the following. "You divided the money in the following way. You: € . Participant B: €. You can now explain ("*begründen*", in German) this decision for yourself." We asked subjects to stick to a maximum of two or three sentences and imposed a generous upper bound of 500 characters.

[14] Additionally, raters evaluated the narratives with regard to their creativity, profoundness, and honesty. We do not use these measures in this study.

random pairs of subjects. One of the 6 choices is randomly selected and the choice of one of the two subjects in the pair is randomly implemented. For further details on the measure, we refer to Murphy et al. (2011).

The SVO measure has been shown to be a stable and consistent predictor of behavior in different social dilemma settings (see Balliet et al., 2009, for a meta-study). Moreover, high SVO types (prosocials) have been shown to differ from low SVO types (selfish) in their decision-making process (e.g., Fiedler et al., 2013). This makes the SVO measure particularly suitable for capturing heterogeneity in reactions to our narrative manipulation.

We additionally elicit further psychological measures. We include the 11-item, Big5 questionnaire (Rammstedt and John, 2007), the Context Dependence and Independence questionnaire (Gollwitzer et al., 2006), a reduced form of the Moral Disengagement questionnaire (Bandura et al., 1996), and a modified version of the Moral Identity Scale (Aquino and Reed, 2002) (for more details on these measures, see Appendix A.1.4). We use these measures (a) as controls in a robustness check in our regression analysis, and (b) to explore the role they play in explaining our treatment effect.

Ex-post measures. Directly after the dictator game decision, subjects went through a series of stages meant to investigate potential mechanisms driving our treatment effects. We describe the questions in the order in which they were presented to participants.[15]

1. General happiness and contentment.
2. Feelings with regard to dictator game choice: happiness, guilt, content, amusement, shame, pride and excitement.

Procedures. The experiment was conducted at the DecisionLab of the Max Planck Institute for Research on Collective Goods in Bonn between May and June 2018.[16] The online experiment was conducted using Qualtrics, while the laboratory experiment was programmed in zTree (Fischbacher, 2007). Subjects were recruited via Orsee (Greiner, 2015). Before the start of the laboratory experiment subjects had to answer control questions to make sure they understood the experimental instructions correctly. 282 participants (64% female, average age 24.8 years)[17] took part in the experiment. For the analysis, we exclude 2 subjects who had not taken part in the online pre-study. All subjects received a show-up fee of 5 €, plus their earnings from the the online pre-study (2 € participation fee plus between 0.50 € and 3 € for the SVO slider task) and their earnings from the dictator game. Overall, subjects received an average payment of 14.48 €. The online pre-study lasted between 5 and 15 minutes, while the laboratory experiment took on average 40 minutes.

[15] We also asked subjects to state their personal norm, i.e., how much they thought would be appropriate to give. However, since the measure was elicited after subjects had made their choice, we cannot exclude that it was used in a self-serving manner to further justify their choice. In fact, we find no variation between treatments and a high correlation with giving. For these reasons, we do not use this measure in our analysis.

[16] For an overview over all sessions, see Appendix A.1.5.

[17] For 74 subjects, this information was not recorded.

1.2.2 Behavioral Predictions

We develop a simple theoretical framework describing how prosocial behavior is influenced by narratives and derive benchmark predictions for the effect of our treatment conditions. Our approach builds on Bénabou et al. (2018), from which we borrow some key notions. While their aim is to study a broad set of phenomena, such as the emergence of narratives and their interpretation or transmission, we focus on getting a deeper understanding of the potentially heterogeneous effects of positive and negative narratives on different social types.[18] This gives us a self-contained theoretical framework for which we provide an intuitive description below (the full version can be found in Appendix A.2). We first outline the reasoning leading up to our hypothesis on aggregate behavior, and then further qualify our predictions for heterogeneous social types.

We start with the notion that decision makers are more inclined to act prosocially the more the consequences of their actions benefit others or the public good (e.g., Goeree et al., 2002, and see the discussion in Bénabou and Tirole, 2006). In turn, this influences the extent to which an action is perceived as appropriate. As the literature on social norms shows, changes in what is perceived as socially appropriate reliably predict changes in behavior across several settings (Krupka and Weber, 2013).[19] Similarly, decision makers care about the deservingness of the recipient(s) of their prosocial action. In distributional choices, decision makers want to avoid giving too much to an undeserving recipient and too little to a deserving recipient (Cappelen et al., 2013). However, the true deservingness of recipients is often unknown in the real world (Cappelen et al., 2018). Likewise, the perception of what is deemed as appropriate is highly flexible and prone to self-serving interpretations (Gino et al., 2016).

Narratives in our setting are arguments targeting these perceptions of deservingness or appropriateness. A positive narrative could, for example, state that the recipient is as deserving as the dictator, because both spent the same time in the lab or because roles were assigned by a random draw. By contrast, a negative narrative might undermine the perceived appropriateness of giving, e.g., by arguing that it is not necessary to give to an anonymous recipient or that everyone else would also behave selfishly, questioning the deservingness of other participants. Importantly, these stories only need to be convincing in the sense of influencing a decision maker's perception of the choice. If positive or negative narratives are indeed successful in changing the perception of the decision maker, they will influence behavior. Our hypothesis on aggregate behavior follows directly.

Hypothesis 1.1 *Positive narratives increase giving, while negative narratives decrease giving.*

We now look at how the perception, and hence the behavior, of different social types is influenced by negative and positive narratives. As mentioned above, the deservingness of a recipient and

[18] In the model by Bénabou et al. (2018), types are defined as either moral or immoral. In our setting, we look at a continuum of types, where heterogeneity stems from diverging beliefs about the appropriateness and the consequences of an action.

[19] The main intuitions we derive from our theoretical framework also hold in a social norms framework with heterogeneous inclinations to follow the norm, as we describe in Appendix A.2.

the appropriateness of giving are subject to uncertainty, and their perception can be influenced by narratives. This uncertainty leaves room for diverging perceptions.[20] In our setting, we call decision makers who perceive a recipient to be deserving or giving as appropriate "prosocial" types, and the ones who believe the opposite "selfish" types.[21]

Consider a prosocial decision maker who hears a negative narrative undermining her perception of the recipients' deservingness. If, as we assume above, she ascribes some truth to the narrative, her perception, and hence her behavior, will change and lead her to give less. Importantly, this effect will be greater compared to that of the same negative narrative on a selfish decision maker, who had a lower perception of the recipients' deservingness in the first place. Vice versa, a positive narrative will have a greater effect on a selfish compared to a prosocial decision maker.

Hypothesis 1.2 *Positive narratives have a stronger effect on more selfish types, while negative narratives have a stronger effect on more prosocial types.*

1.3 Results

Our dataset consists of 280 independent observations spread over three experimental conditions. In the first part of this section, we analyze the evidence regarding our main hypotheses. We then provide additional insights on the way our treatment conditions influence behavioral results.

1.3.1 Main results

Subjects in the Baseline condition give on average 2.76 €. According to Hypothesis 1.1, we should observe an increase in average giving in the Positive condition and a decrease in the Negative condition. Figure 1.2 provides a visual representation of the aggregate results. In the Positive condition, average giving increases to 3.23 €. This constitutes a 17% increase, in line with our first hypothesis. The difference, however, is only marginally significant (rank-sum test, $p = .093$). Average giving in the Negative condition (2.78 €) is virtually identical to average giving in the Baseline condition (rank-sum test, $p = .908$).[22]

However, the aggregate results on giving provide an incomplete picture of the data. As stated in Hypothesis 1.2, prosocial types should respond more strongly to the Negative treatment condition and selfish types to the Positive treatment condition. Although the effect should go in the same direction for all types.

Figure 1.3 displays the relationship between how much a subject gave in the dictator game and her social type. Giving is, as is typical in dictator games, bounded above at 5 € with only two

[20] We are agnostic about where these different perceptions come from and simply require them to influence behavior. They may be deeply grounded in a decision maker or may have formed through experience, or else a decision maker might self-servingly hold a perception which allows her to act in a certain way.

[21] In our experiment, we use the Social Value Orientation to measure these different perceptions. A higher (lower) SVO angle corresponds to a higher (lower) perception of deservingness or appropriateness.

[22] The difference in giving between Positive and Negative is not significant (rank-sum test, $p = 0.114$).

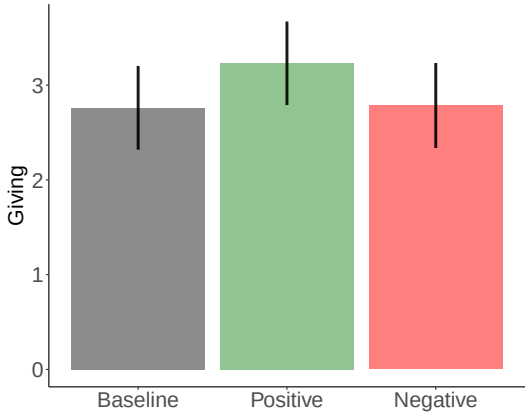


Figure 1.2: Average giving with 95%-confidence intervals.

subjects giving 6 € and many giving nothing at all. We use LOESS fitted lines to provide a better visualization of the data. The black solid line depicts the relationship between the social type and giving in BASELINE; the green dotted line represents our POSITIVE condition and the red dashed line our NEGATIVE condition. We observe the expected positive correlation between our social type measure and giving in the BASELINE condition. The steepness of the fitted line in the middle of the graph indicates that, in line with previous studies (see Engel, 2011), giving follows a bimodal distribution, with many subjects giving either half of their endowment or nothing at all.

dv: giving	(1)	(2)
POSITIVE	0.752**	2.852***
	(0.360)	(0.888)
NEGATIVE	0.125	2.698***
	(0.360)	(0.894)
Type	0.133***	0.189***
	(0.0116)	(0.0217)
POSITIVE x type		-0.0732**
		(0.0283)
NEGATIVE x type		-0.0900***
		(0.0285)
Constant	-1.382***	-3.015***
	(0.428)	(0.696)
Observations	280	280
Pseudo R^2	0.108	0.118

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 1.1: Tobit regressions.

Note: Coefficients of Tobit regression with lower censoring at 0. The type measure corresponds to the SVO angle, POSITIVE and NEGATIVE conditions are introduced as dummies. We also include interaction terms between conditions and the SVO angle in column (2).

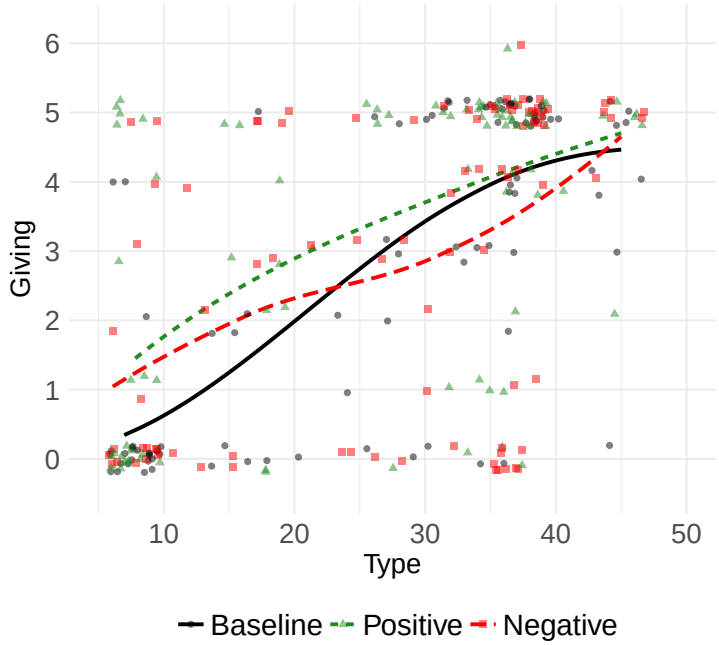


Figure 1.3: Giving on SVO. LOESS fitted lines.

Note: Data points are jittered. For the ease of visualization, we removed social types below 5° and above 50°, which are rare (5 subjects) and not balanced across treatments.

To test how different types react to different narratives, we run a Tobit regression with the amount of giving as the dependent variable and treatment dummies, type, and interaction terms between type and treatment dummies as explanatory variables (see Table 1.1). However, subjects' SVO-angles are not distributed uniformly (see Figure 1.4).[23] The modal selfish type (60 subjects with an SVO angle of 7.82°) and the modal prosocial type (61 subjects with an SVO angle of 37.48°) make up 43% of all observations. Thus, we also look at them in isolation to complement the regression analysis and provide a sanity check for our results. We discuss further robustness checks at the end of this section.

We first look at column (1), where we introduce our treatment conditions as dummies and control for the social type of a subject. The Positive condition has a strong positive and significant effect on giving, confirming part of Hypothesis 1.1. The overall effect of the Negative condition is also positive, but small and not significant. Note that, as expected, the type measure is a clear predictor of giving: the higher the SVO angle of a subject, the more she gives.

In column (2) we add an interaction between subjects' social type and the treatment conditions. To interpret these results we plot the estimated marginal effects of our treatment conditions on giving compared to the Baseline in Figure 1.4. This enables us to test Hypothesis 1.2.

We start with the Positive condition (green dotted line), where we find a pattern in line with our hypothesis. We notice a strong positive effect for more selfish types, which fades out for

[23] Due to the construction of the measure specific SVO angles appear more frequently in the data (see Murphy et al., 2011).

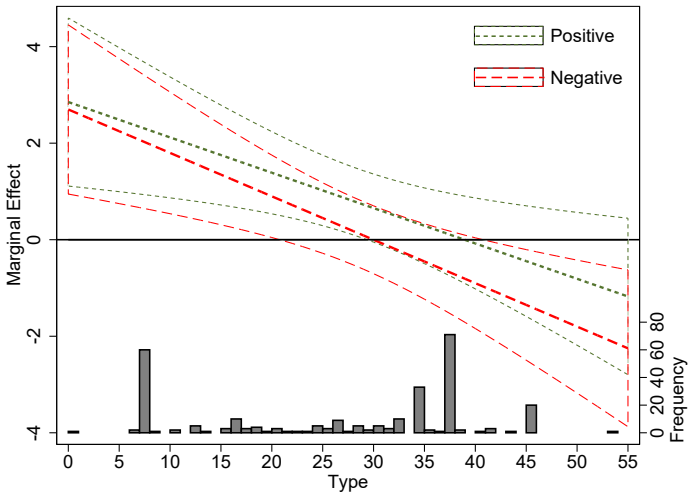


Figure 1.4: Marginal effects on types, 95% confidence intervals.

Note: In the lower part of the graph, we plot the pooled distribution of types over all conditions. Numbers indicate the SVO-angle with higher angles indicating more prosociality. For the ease of visualization, types below 0° (3 subjects) are not displayed.

more prosocial types. The estimated marginal effect for the modal selfish type corresponds to a positive and significant difference of 2.28 € ($p = .001$) in giving, compared to the Baseline. Prosocial types, on the other hand, display no significant increase. This finding is corroborated by comparing giving in the Positive condition with the Baseline for the modal selfish (t-test, $N = 46$, $p = .028$) and prosocial types (t-test, $N = 39$, $p = .770$) in isolation.

Result 1.1 (Positive Narratives) *Positive narratives increase giving compared to the* Baseline *condition. This effect is driven by more selfish types.*

In the Negative condition (red dashed line), more selfish types increase their giving compared to the Baseline. The estimated marginal difference of 2 € ($p = .004$) for the modal selfish type is positive and significant. Note that this increase is indistinguishable from the one of the Positive condition. This is clearly not in line with our hypotheses. More prosocial types, on the other hand, give less than in the Baseline. The modal prosocial type decreases giving by an estimated marginal difference of 0.67 € ($p = .121$), which is not statistically significant. However, for more prosocial types (21 subjects with an SVO angle above 44°), the effect becomes negative and significant. These results are confirmed when restricting the analysis to the modal selfish type (t-test, $N = 37$, $p = .030$) and modal prosocial type (t-test, $N = 42$, $p = .016$), who increase and decrease giving, respectively.[24]

Result 1.2 (Negative Narratives) *Negative narratives have a differential effect: they decrease giving for more prosocial types and increase giving for selfish types compared to the* Baseline.

We run further regressions to check the robustness of our results (see Appendix A.3). First,

[24]Note that this is in line with the LOESS fit presented in Figure 1.3.

we compare the results from the Tobit regressions with a standard OLS regression. We then include the additional psychological measures collected in the online pre-study and session dummies as controls in our Tobit model. We also run a Tobit model with both lower and upper censoring. Finally, we include a quadratic interaction term between our treatment conditions and the social type to capture potential nonlinearities. Our results are robust to these additional analyses.[25]

1.3.2 Additional results: do people follow the narrative?

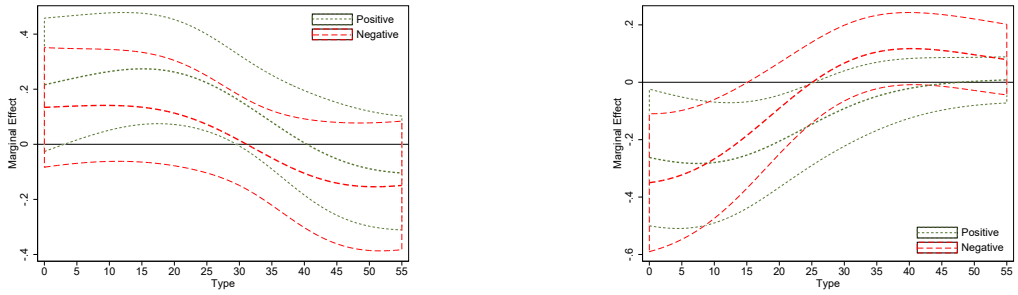


Figure 1.5: Marginal effects, Probit

Note: The dependent variable is a dummy for giving 5 € on the left and for giving 0 € on the right. Explanatory variables are: the SVO angle, dummies for the Positive and the Negative condition and interaction terms between treatment conditions and the SVO angle. Outer lines show 95 % confidence intervals. For the ease of visualization, subjects with an SVO angle below 0° (3 subjects) are not displayed.

A natural question is whether narratives led subjects to adhere to the behavioral prescription contained in them, i.e., either to share equally or keep everything for themselves. In other words, did the Positive (Negative) condition lead subjects to give 5 € (0 €) more frequently than in the Baseline?

To answer this question, we run two Probit regressions on the probability of giving either 5 or 0. The graphs in Figure 1.5 show the estimated marginal effects on different social types for the same specification we used in our main regression in Table 1.1 column (2) (see Table A4 in Appendix A.3.2 for the full regression results). There are three main observations to be made. First, the left graph in Figure 1.5 shows that the probability of giving an amount equal to 5 € in the Positive condition increases for nearly all selfish types.[26] This translates into a 26% higher probability of giving 5 € for the modal selfish type in the Positive condition (estimated marginal effect, $p = .022$). In the Negative condition, on the contrary, the increase in the probability of giving 5 € is smaller and statistically insignificant. The difference for the modal selfish type is just 14% and not significant ($p = .178$). Second, the right graph in

[25] We also perform our analysis using the Moral Identity Scale and the Moral Disengagement questionnaire as alternatives to the SVO angle in our main regression. Both have a strong and stable relationship with giving, but turn out to be irrelevant in explaining our treatment difference. Moreover, Context Dependence or Independence do not mediate our treatment effects. This gives us further assurance in using the SVO as our type measure for the main analysis (see Appendix A.3.1) for further details.

[26] The effect is particularly strong for the range of selfish types who are more frequent in our sample (those above an SVO angle of 5° and below one of 25°).

Figure 1.5 shows that both in the POSITIVE and the NEGATIVE condition the probability of selfish types giving 0 decreases substantially. This effect is observed across a wider range in the POSITIVE condition. The estimated marginal decrease in the probability of giving 0 € for the modal selfish type corresponds to 28% ($p = .012$) and 30% ($p = .007$) in the POSITIVE and NEGATIVE condition, respectively. Third, we find that, although more prosocial types give less in the NEGATIVE condition, this does not lead to a substantial increase in the probability of giving 0 €. The increase in probability for the modal prosocial type is moderate (11%) and only marginally significant ($p = .077$).

Result 1.3 *The* POSITIVE *condition increases the probability of giving* 5 € *for selfish types. Both treatment conditions decrease the probability of selfish types giving* 0 €*.*

We finally look at the effect of our treatment conditions on the ex-post measures of subjects' feelings (Table A5 in Appendix A.3.3 shows our regression analysis). We find no treatment effects on general happiness or contentment. Feelings of guilt and shame with regard to the choices made by subjects have, as one could expect, a strong and stable relation with the amount of giving: giving less increases these reported feelings. However, our treatment conditions do not increase or reduce guilt or shame about choices. Nevertheless, we cannot rule out that the absence of treatment effects is caused by the anticipation of these feelings. The presence of narratives could lead subjects to anticipate guilt or shame and to adapt their giving to avoid them, which could result in similar stated feelings across treatments.

Result 1.4 *Our treatment conditions do not directly change subjects' feelings towards their choice.*

1.4 Discussion and Conclusion

Our results provide insights into how narratives in favor of prosocial or selfish actions influence the behavior of different social types. Subjects in our experiment see either positive or negative narratives upon taking a distributional choice in a dictator game. We compare our two treatment conditions with a baseline in which no narratives are provided. Empirical beliefs about the distribution of choices are fixed across all experimental conditions. We work out two hypotheses from a theoretical framework on how narratives influence behavior via the perception of the appropriateness of an action or the deservingness of a recipient for different social types.

Subjects in the POSITIVE condition give more than subjects in the BASELINE condition. This increase is predominantly driven by selfish types (Result 1.1). On the other hand, narratives in the NEGATIVE condition have a differential effect (Result 1.2). Prosocial types in the NEGATIVE condition give less than in the BASELINE. However, this effect is reversed for selfish types, who give more in the NEGATIVE condition compared to the BASELINE, matching the giving level of their peers in the POSITIVE condition. These results are only partly in line with the hypotheses derived from our theoretical framework. In particular, our hypotheses allow the effect of narratives to have different strength for different social types, but predict that all social

types should move in the same direction. This suggests that narratives could have an effect beyond that of arguing in favor or against the appropriateness of a certain action, as we describe below.

The differential effect of narratives resonates well with other research showing that different social types process information differently (Fiedler et al., 2013) and have a different representation of moral dilemmas (Van Lange et al., 1990; Liebrand et al., 1986). This suggests that our manipulation of positive and negative narratives could, indeed, affect prosocial and selfish types differently. We suggest two potential explanations which can account for the differential effect: one based on the argument that narratives enhance the moral saliency of the decision and another one based on a social comparison motive.

According to the first explanation, as pointed out above, the more selfish individuals might disregard the consequences of their actions and of the presence of a norm in their "ordinary" decision process. They could genuinely not know or deceive themselves. In both cases, the mere presence of a narrative, regardless of its content, could make the moral nature of the situation and, hence, the norm more salient, leading selfish individuals to give more. This conjecture is in line with a study by Krupka and Weber (2009), who find that descriptive information enhances prosocial behavior, even in cases where one does not observe a lot of norm-compliant behavior. Similarly, Gino et al. (2009) find that increasing the saliency of an opportunity to cheat decreases unethical behavior. This also resonates with a study by Xiao (2017) who shows that the pressure to justify leads to more norm-compliant behavior in prosocial choices. In this sense, the moral salience induced by narratives might lead "reluctant sharers" to give (Lazear et al., 2012). An account based on this moral saliency effect, however, does not explain why prosocial types decrease their giving when faced with a negative narrative, since the norm should be salient for them as well.

Our second explanation based on social comparison, instead, can account for the whole pattern of our main results. If subjects care about how they fare in the comparison with others, the content of the narrative could serve as a social benchmark. In particular, narratives in the Negative condition would represent a very low reference point. Giving at least something after facing a negative narrative provides a low-cost opportunity for a selfish type to distinguish herself from the narrator. At the same time, prosocial types are led to give less by the negative narrative, but still care about faring well in the comparison with the narrator. In the Positive condition, on the other hand, the social benchmark is set very high. For a subject not to look bad in this comparison she has to match the giving of the narrator. Taken together, this would mean that subjects want to distinguish themselves from a selfish narrator and imitate a prosocial narrator. In Appendix A.2.1, we extend our model by including a social comparison component and provide a specification that can rationalize our results.

The additional results we obtain from our Probit regressions with either the equal split or the selfish action as dependent variables (Result 1.3) further support the social comparison explanation. The observed increase of equal splits in the Positive condition suggests that at least

some subjects wanted to avoid the negative comparison with the prosocial narrator and imitated her behavior. In the NEGATIVE condition, the probability of giving nothing decreases for selfish types and does not increase for prosocial types, implying that subjects were driven by a desire to differentiate themselves from the selfish narrator at least marginally. This behavior is in line with the phenomenon of partial lying (Fischbacher and Föllmi-Heusi, 2013) or ethical maneuvering (Mazar et al., 2008; Shalvi et al., 2011b), which is consistently found in the experimental literature on lying and cheating. Subjects often do not lie to the full extent, in order to avoid being unequivocally identified as liars or cheaters. This motivation is very similar to that of prosocial subjects in our experiment who only marginally decrease their giving.

Note that a social comparison explanation does not contradict the above point that narratives heighten the normative salience of the decision. Neither does it go against the evidence cited to support that explanation. Far from it, we in fact argue that narratives evoke a salient, vivid benchmark subjects compare themselves with. This account is supported by psychological theories which emphasize the importance of social comparison for people's self-evaluation (see the seminal work of Festinger (1954) and Suls and Wheeler (2013) for an overview) and its crucial role for normative behavior (Cialdini et al., 1990, 2006). Social comparison has found fertile ground in economics as well and has sparked research in many different areas from energy and water conservation behavior (Allcott and Rogers, 2014; Ferraro and Price, 2013), to public good provision (Shang and Croson, 2009), charitable giving (Frey and Meier, 2004), all the way to retirement savings decisions (Beshears et al., 2015).

Importantly, our study was not designed to specifically test the social comparison mechanism. Psychological theories (Tesser, 1985) emphasize that for a comparison to be meaningful for an individual, she has to feel close to the person she is comparing herself with. In other words, the comparison has to be self-relevant. In our experiment, we did not manipulate the relevance of the comparison with the narrator. This could be done, e.g., by choosing narrators that either belong to the same social category of the subject or to a different one. Since we use a student sample and subjects in our sample are used to face other students in these experiment, there are good reasons to believe that the comparison was relevant for them.

Our work advances the understanding of the determinants of prosocial and moral behavior by providing insights into how narratives - which permeate people's life - work. Our findings suggest that narratives sway subjects while, at the same time, serving as a benchmark for social comparison. Arguments in favor of selfish or prosocial behavior seem to evoke a concrete normative dilemma in subjects' mind. To be or not to be like the narrator? How will I fare compared to her? Subjects react to this vivid image by adhering to the narrative of a prosocial narrator and wanting to distinguish themselves from a selfish narrator. Certainly, more research is needed to understand how exactly this process works.

Our results also have relevant implications for institutions and organizations who can use narratives to promote prosocial behavior, especially amongst the people who would be less inclined to act so ex ante. This can be achieved by confronting people with different narratives, positive or

negative, depending on their predisposition. In the setting we study, sharing the money equally represents a clear norm of behavior. Future research could investigate the relationship between narratives and the strength of a norm or the presence of multiple norms. Other questions are how enduring the effect of a certain narrative is, and whether there might be spillovers in other contexts. We hope our work can contribute to inspire such endeavors.

Chapter 2

Upset but (almost) correct: A robustness check of di Tella, Perez-Truglia, Babino and Sigman (2015)

In recent years, a large body of literature has researched the conditions under which people display moral behavior and those under which they act in their own self-interest. These different strands of literature highlight the human need to look for motives when faced with moral dilemmas (Shafir et al., 1997; Ditto et al., 2009) to help reduce cognitive dissonance (Festinger, 1962). People often develop biased views about the facts that surround them and about the people they interact with to resolve the tension between being or appearing moral and pure self-interest (for some examples see Miller and Ross, 1975; Kahan et al., 2012; Gino et al., 2016). Examples of such self-serving beliefs include self-serving manipulation of fairness arguments (Konow, 2000), of ambiguity (Haisley and Weber, 2010) and of risk (Exley, 2015).

In a recent paper, Di Tella et al. (2015) investigate the formation of self-serving beliefs justifying unfair behavior in a modified Dictator Game, the "corruption game". In the "basic" version of the game, an Allocator decides how to split 20 tokens between herself and a Seller. The Seller decides how to convert the tokens into money. She can decide to convert the money at a high conversion rate (1 token = 2 AR$ (Argentinian Pesos)) or a lower one (1 token = 1 AR$), in which case she receives a side-payment or "bribe' (10 AR$). Decisions are taken simultaneously. Importantly, the Allocator is asked to state her beliefs regarding the behavior of the Seller she is paired with and that of Sellers in her session. The former beliefs are not incentivized, while the latter are. A self-serving bias in beliefs can be identified by comparing the two between-subjects treatments of Di Tella et al. (2015). In both of their treatments, 10 of the 20 tokens are assigned to the Allocator and 10 to the Seller. In the *Able=8* treatment, an Allocator can give or take up to 8 tokens. This means she can choose any allocation ranging from 2 tokens for herself and 18

for the Seller, to 18 tokens for herself and 2 for the Seller. In the *Able=2* treatment, the ability to take is restricted, and an Allocator can give or take only up to 2 tokens. Hence, in this case, she can choose any allocation ranging from 8 tokens for herself and 12 for the Seller, to 12 tokens for herself and 8 for the Seller.

The key ingredient of this design is that Sellers do not know about these restrictions and believe that Allocators can give or take any amount they want in both treatments. In turn, Allocators know that the Sellers are not aware of any restriction. Hence, according to Di Tella et al. (2015), any treatment variation cannot be ascribed to changes in the expectations of Allocators with respect to Sellers' behavior and must be due to internal reasons. In particular, an Allocator who cares only about her material consumption has no reason to manipulate her beliefs about the Seller. She will state her true belief as she is incentivized to do. However, an Allocator who also cares about her self-image (Bénabou and Tirole, 2006) might engage in self-deception and justify allocating more tokens to herself in the *Able=8* by self-servingly believing that the Seller chose to take the bribe. The authors also develop a model capturing this intuition via a reciprocity mechanism. In the comparative statics of their model beliefs and behavior move in the same direction, i.e., if an Allocator takes more she does so by thinking that more Sellers chose the bribe, i.e., that the probability to be paired with a corrupt Seller is higher. The findings of the authors are in line with this interpretation. Indeed, Allocators in the *Able=8* treatment think that more Sellers chose the unfair option, accepting the side-payment and leaving them with less money.

This robustness check of the basic corruption game of Di Tella et al. (2015) changes minor details in the design, but fails to reproduce the main findings. Allocators' beliefs, in this experiment, are not self-servingly biased. If anything, Allocators in the *Able=8* treatment are less upset than those in the *Able=2* treatment. The results of this experiment are also discussed in relation to a recent paper by Ging-Jehli et al. (2019), who study the role of self-serving strategic beliefs in a "pre-emptive taking game" and also replicate the "modified" corruption game of Di Tella et al. (2015).

2.1 Experimental Design

The basic corruption game of Di Tella et al. (2015) was run with few changes in the design to maintain the basic structure of the experiment. Instructions were translated into German, but, differently from the original study, the use of loaded language was avoided (see Appendix B.1). The Allocator and the Seller were simply called participant A and B. The two options the Seller could choose from were simply called Option 1 and Option 2. On the contrary, in the original instructions, Allocators were told that the experimenter would have preferred the Seller to take the bribe. Note that this was not the way the options were described to the Seller herself and that Allocators saw their instructions too. Moreover, a different real effort task was used at the

beginning of the experiment.[1] As in the original experiment the task lasted for around 1 minute and participants earned 10 tokens by completing it.

The main task of the experiment remained substantively unchanged. The Allocator could choose how to distribute the 20 tokens between herself and the Seller. Her choice was restricted depending on the treatment. In the *Able=8* treatment, she could give or take up to 8 tokens. In the *Able=2* treatment, she could give or take only up to 2 tokens. The Seller simultaneously decided how to convert the tokens into Euros. Each token, was worth 0.50 € if she chose the high conversion rate. If she chose the low conversion rate, instead, each token was only worth 0.25 €, but she got an additional payment of 5 € just for herself. After the main task and before knowing the actual outcome of the interaction, the beliefs of the Allocator were elicited as detailed below.

Apart from the minor adaptations mentioned above, two more further changes were made. First, Allocators were told from the beginning that they could give or take only 2 or 8 tokens, depending on the treatment they were in. On the contrary, in the original design, Allocators learned about this fact only upon taking their allocation decision. This design choice differs also from the modified version of the corruption game that the authors use as a robustness check in their paper. There Allocators were told about the existence of the both treatments and that they had been assigned to one of the two. The second substantive change with respect to the original design, consists in the way incentivized beliefs were elicited. As in the original study, Allocators were asked whether they thought that the Seller they were paired with took the bribe. As to the incentivized measure, Allocators in the original study were asked to estimate the percentage of Sellers who chose the unfair option in their session and were given 10 predefined brackets going from 0-10% to 90-100%. Subjects were rewarded if the actual percentage of Sellers choosing the unfair option fell in the bracket they had chosen. While it may be easy to understand, this method has one major drawback. Depending on the number of subjects in a session, a single bracket could contain no correct response at all or more than one, making the brackets unequally likely and potential distorting incentives in the elicitation procedure. This introduces noise in the beliefs data. Take, e.g., the case with 12 Sellers. If an Allocator puts the highest probability, say 30%, on there being 8 out of 12 Sellers (66%) who took the bribe, she should go for the bracket 60-70%. However, say that this Allocator also believes that there is a 20% probability on there being either 6 or 7 out of the 12 Sellers who chose the unfair option. This means that she assigns a 40% probability on the 50-60% bracket and will thus select it. This yields a biased estimate of the true beliefs. Similarly, if there are only a total of 8 Sellers, some brackets would be empty, which will also distort incentives. To avoid this, Allocators were told the actual number of Sellers present in their session and were asked to estimate the exact number of Sellers who chose the unfair option. With at most 16 participants per session and hence 8 Sellers, this means they had to state a number between 0 and 8. The shift from percentages to natural numbers should not interfere with the formation of self-servingly biased beliefs, which Allocators are likely to have

[1] In the task used by the authors, participants had to find a given hidden sequence of 0s and 1s in a series of 0s and 1s. Participants in this experiment, instead, had to complete a computerized slider task (see Appendix B.1).

formed already before they reach the beliefs elicitation stage. Correct beliefs in the incentivized elicitation were rewarded with 3 €.

2.1.1 Procedure

The experiment was run at the Cologne Laboratory for Economic Research (CLER) during June and July 2018. The entire experiment was programmed using z-Tree (Fischbacher, 2007). Participants were recruited via ORSEE (Greiner, 2015). A total of 118 subjects (55% female, average age 24.9 years) took part in the experiment across 8 sessions. This yields 29 independent observations in the *Able=2* and 30 in the *Able=8* treatment. Based on the effect size of the basic corruption game in Di Tella et al. (2015), 30 observations per treatment were needed to achieve a power of 80%.

Subjects received on screen instructions. After reading the instructions they went through a series of control questions. Once they answered all questions subjects received personalized, on screen, feedback about their answers. Only after all participants had finished this phase, the actual experiment started. The experiment lasted between 30 and 40 minutes overall. Subjects earned 9.74 € on average, including a show-up fee of 4 €.

2.2 Results

	Basic Game (di Tella et al., 2015)			This study		
	Able=2	Able=8	p-value	Able=2	Able=8	p-value
Tokens taken	1.53	6.00	<0.01	1.14	4.07	<0.01
	(0.22)	(0.72)		(0.18)	(0.76)	
Is Corrupt	0.47	0.87	0.02	0.86	0.57	0.02
	(0.13)	(0.09)		(0.06)	(0.09)	
%-Corrupt	0.49	0.69	0.08	0.76	0.65	0.11
	(0.09)	(0.06)		(0.04)	(0.05)	
N	15	15		29	30	

Table 2.1: Comparison of Allocators' behavior

Note: Average characteristics with standard errors in parentheses. P-values of standard mean difference test with null hypothesis that means are equal in Able=2 and Able=8. The measure of %-Corrupt for this study is constructed by transforming the results from the incentivized belief elicitation in percentages.

The first part of this section compares the results of this study with those of the original study. The second part digs deeper in the patterns observed to understand what drives them.

Table 2.1 offers a comparison between the original study and the present one, reproducing the main analysis presented by the authors in their paper. It compares Allocators' behavior in the *Able=2* and *Able=8* treatment. Allocators in the present study took slightly less tokens in both treatments compared to the original study. More strikingly, there is a failure to replicate the pattern of results concerning Allocators' beliefs about Sellers. The difference between the two

treatment conditions for the unincentivized measure (*Is Corrupt*) goes in the opposite direction and is highly significant. 86% of Allocators in the *Able=2* treatment think that the Seller they were paired with chose the unfair option, while only 57% of Allocators in the *Able=8* treatment believes so. The same pattern is also present in the incentivized measure (*%-Corrupt*), but the difference is not statistically significant.

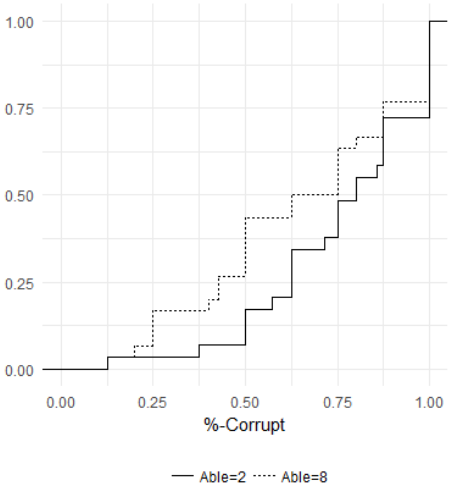


Figure 2.1: Ecdf plot of *%-Corrupt*

Figure 2.1 displays the cumulative density function of *%-Corrupt* in the present study. The solid line corresponding to the *Able=2* treatment lies always to the right of the dotted line of the *Able=8* treatment, meaning that the whole distribution of the incentivized beliefs measure is shifted to the right in the *Able=2* treatment compared to the *Able=8* treatment (Chi-squared test with simulated p-value, $p = 0.01$). This confirms the observations made for Table 2.1 showing that the shift in beliefs is substantial and goes in the opposite direction compared to the original study.

(a) Equalizers **(b)** Takers

Figure 2.2: Treatment comparison of *%-Corrupt* for Equalizers and Takers for this study. Average beliefs with standard errors.

To dig deeper in the data, Figure 2.2, compares the levels of %-Corrupt between the *Able=2* and *Able=8* treatment for two sub-samples of Allocators. In the left chart only Allocators who decided to split the tokens equally are considered, while on the right all Allocators who took at least one token are included. "Takers" hold very similar beliefs across the two treatments

($p = 0.61$). On the other hand, there is a significant difference in the beliefs of "Equalizers" ($p = 0.01$). Allocators who split the pie equally in the *Able=2* treatment are much more upset than their peers in the *Able=8* treatment. Interestingly, the percentage of Sellers who actually choose the unfair option in this experiment is 80%, which almost perfectly coincides with the average belief of takers (81%).

(a) Equalizers **(b)** Takers

Figure 2.3: Treatment comparison of *%-Corrupt* for Equalizers and Takers for Di Tella et al. (2015). Average beliefs with standard errors.

Figure 2.3, reconstructs the same comparison with the original data from (Di Tella et al., 2015). As in the present study there is no significant difference in Takers' beliefs ($p = 0.28$). A similar test for Equalizers cannot be performed, since there is only one observation in the *Able=8* treatment. However, testing against the value of this observation (0.7) yields a marginally significant result ($p = 0.08$). This means that, while Equalizers in the present study had more negative beliefs in the *Able=2* compared to the *Able=8* treatment, the data in Di Tella et al. (2015) go in the opposite direction. In both cases the results are driven by the beliefs of Equalizers.

2.3 Discussion and Conclusion

The robustness check of Di Tella et al. (2015) presented in this chapter failed to replicate their results. If anything, the unincentivized beliefs measure (*Is Corrupt*) suggests that Allocators who were able to take only little in the *Able=2* treatment were more upset than those who could take more in the *Able=8* treatment. A comparison between the distributions of the incentivized beliefs measure (*%-Corrupt*) in the *Able=2* and *Able=8* treatment goes in the same direction. Moreover, beliefs in the *Able=2* treatment are objectively less biased, since they are closer to the actual share of Sellers who choose the unfair option (80 %).

A comparison of Allocators who split the pie equally with those who take something from the Seller, shows that Equalizers in the *Able=2* treatment hold much more negative beliefs than their peers in the *Able=8* treatment. Taken together this evidence seems to suggest that Allocators who equalize the tokens are rather optimistic in their beliefs about Sellers. In turn, they become more "realistic" when they take. In this sense, they are upset, but almost correct in believing that their counterpart misbehaved. Interestingly, this is true also for the Allocators who split the tokens equally in the *Able=2* treatment.

There were three main changes undertaken compared to the original study. First, the instruc-

tions used a neutral language without any framing. Allocators and Sellers where simply called Participant 1 and 2. The bribe was not described as the preferred option by the experimenter. Second, beliefs for *%-Corrupt* were elicited in a different, less noisy way. Finally, Allocators were told at the beginning of the experiment to which treatment they had been assigned without mentioning the presence of another treatment. This changes should not alter the nature of the strategic interaction in the experiment or interact with the formation of self-serving beliefs.

A potential reason behind the failure to replicate the results from Di Tella et al. (2015) is the very negative (although almost correct) level of beliefs of Allocators in the present experiment. A large share of Allocators thought that that most sellers would take the bribe. This means that the tokens were worth very little for them. If Allocators care about their self-image and if taking anything away from Sellers results in the breach of a norm, thus hurting their self-image, Allocators could choose to withhold a positive self-image by not taking anything. Allocators in the *Able=8* treatment could compensate a loss in self-image by taking more tokens away from the Seller, which would almost equalize their payoffs. Allocators in the *Able=2* treatment, instead, could only take very little, which makes the equal split more attractive for them. Hence, some of the Allocators who would have otherwise chosen to take decided not to do so in the *Able 2* treatment. As a consequence, the beliefs of Equalizers in this treatment become more negative, because they include both genuine Equalizers and Takers, who decided to forego taking to withhold a positive self-image.

In a recent paper, Ging-Jehli et al. (2019) study the role of self-serving strategic beliefs in a "pre-emptive taking game" and also replicate the "modified" corruption game of Di Tella et al. (2015). In the pre-emptive taking game, two subjects start off with the same endowment and a first-mover can decide to take from the second-mover, thus limiting the second-mover's ability to take from her. In this game, Ging-Jehli et al. (2019) find no evidence of self-serving beliefs by comparing the beliefs of the first-mover and an unbiased third party about the second-mover.

The modified corruption game that Ging-Jehli et al. (2019) replicate differs from the basic version studied in this chapter in various ways. The most important difference is in the options available to the Seller. While the corrupt option is a weakly dominant strategy in the basic corruption game (provided Allocators only take from and do not give to the Seller), in its modified version the optimal action of the Seller depends on how much the Allocator takes from her. Ging-Jehli et al. (2019) replicate the same pattern of results found in the original paper. Allocators in the *Able=2* treatment believe 33% of Sellers to be corrupt, while the share increases to 47% in the *Able=8* treatment. However, they also elicit beliefs about the Sellers' behavior from a third party. In line with their results on the "pre-emptive taking game", they find no difference between the beliefs of Allocators in the *Able=8* treatment which are supposed to be biased self-servingly and those of third parties. Third parties believe that 46% of Sellers take the bribe. Moreover, these beliefs are closer than those of Allocators in the *Able=2* treatment to the true share of corrupt Sellers (42%).

Ging-Jehli et al. (2019) interpret this as evidence against the presence of an absolute self-serving

bias in Sellers' beliefs in the *Able=8* treatment. They, instead, argue in favor of a "positivity" bias in Allocators' beliefs in the *Able=2* treatment. Such optimistic beliefs can be explained by the theoretical framework of Di Tella et al. (2015) which simply predicts that those who take will have more negative beliefs compared to those who do not take. In fact, Ging-Jehli et al. (2019) also propose a model of reciprocity that explains why Allocators with a limited ability to take in *Able=2* might develop optimistic beliefs.

The results of the present study do not contradict this interpretation. In fact, Allocators who take have very negative beliefs about Sellers. However, given that beliefs in the present experiment are very negative overall, the positivity bias seems to be offset by potential self-image concerns. This suggests that the level of beliefs in the population of Allocators might interact with the manipulation of their opportunity to take. In conclusion, an accurate assessment of a self-serving bias in beliefs should be performed by comparing different benchmarks and considering reasons beyond the manipulation of beliefs to justify one's actions. Future studies should carefully consider these caveats.

Chapter 3

Social norms, personal norms and image concerns

Social norms are one of the most common rationales used to explain behavior that deviates from standard economic models. For decades, they have been used to account for various phenomena, such as prosocial behavior (Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009; Krupka and Weber, 2013; Bénabou et al., 2018), lying aversion (Gächter and Schulz, 2016; Abeler et al., 2019), costly punishment (Fehr and Gächter, 2000; Fehr and Fischbacher, 2004), various labor market outcomes (Fehr et al., 1998; Lindbeck et al., 1999; Akerlof, 1980) and certain dimensions of household behavior (Lindbeck, 1997).

In this chapter, we direct our attention towards social norm's privately held counterpart - personal norms. In contrast to social norms, personal norms are one's personal beliefs about the appropriateness of a certain behavior in a given situation, *irrespective* of society's view. For decades, scientists in neighbouring fields have argued that personal norms are a relevant driver of behavior (see, e.g., Schwartz, 1973, 1977; Cialdini et al., 1991; Bicchieri, 2005), yet - in contrast to social norms - they have been largely neglected in the field of economics. The two norms, however, can strongly differ in many economic contexts. For example, someone who disapproves of wealth redistribution might have a different personal norm about tax avoidance compared to the social norm of the social-welfare oriented society she lives in. Likewise, the normative beliefs of a person who discriminates towards members of other ethnic or socio-economic groups can be in conflict with those of her society that openly fights such discrimination. Similar discrepancies can exist for all economic behavior that is governed by social norms, such as cooperation, trust or honesty. If people care about personal and not only social norms, they are confronted with two (potentially conflicting) normative principles that can determine behavior.

Here, we propose a simple utility framework and design a novel experiment to demonstrate that personal norms are a *strong predictor* of economic behavior. Our findings show they are: i) complementary to social norms in predicting behavior, ii) robust to an exogenous increase in the salience of social norms, and iii) inherently distinct from social norms across a wide range of

economic contexts.

We start by presenting a simple utility framework where people care about their monetary payoff, social norms and personal norms. More precisely, we assume people care about the money they earn from an action, the degree to which this action complies with their beliefs about injunctive social norms, i.e., what society finds appropriate, and the degree to which this action complies with their own private belief about what is appropriate. This captures a decision making process where two normative principles — one imposed from within the person and the other from the society — are decisive for behavior. We then design a novel two-part experiment which allows us to investigate the predictive value of personal norms as well as social norms across four economic games.

In the first part of the experiment, we elicit both social and personal norms in an online survey for four games: Dictator game, Dictator game with tax, Ultimatum game and Third-party punishment game. To do so, we design a simple method to elicit beliefs about personal and social norms with a symmetric procedure. Subjects go through an adapted version of the Krupka and Weber (2013) social norms elicitation, and a symmetric procedure for eliciting personal norms in a randomized order. The main difference between the two procedures is the following: subjects evaluate all possible actions i) "according to the opinion of the society and independently from their own opinion" for social norms, and ii) "according to their own opinion and independently from the opinion of others" for personal norms. We demonstrate that the two norms elicited with this procedure are correlated, but that there is substantial heterogeneity at the individual level across all four games.

In the second part of our experiment, we invite the same subjects to the lab approximately four weeks after the norm elicitation took place. In the lab, they play the four games we elicited the norms from. Importantly, this was not revealed to them prior to the lab experiment. We then connect subjects' behavior elicited in the lab to their beliefs about personal and social norms elicited in the online experiment. We estimate our utility framework by using a conditional (fixed-effect) logit choice model, and show that personal norms — while taking social norms and monetary payoffs into account — are highly predictive of individuals' behavior. This finding holds across all four games individually, as well as when analyzing them together. Our results further reveal that social norms are also strong predictors of behavior, but their predictiveness seems to vary across games. Having demonstrated the strong relation between personal norms and economic behavior in a treatment where decisions remain private (Private), we then analyze the results of a treatment in which we *exogenously* increase the salience of social norms (Social). As economic decisions are rarely taken in a social vacuum, this allows us to investigate the predictive value of the two norms when subjects' actions are observable. Following the reasoning of Bicchieri (2005), we hypothesize that this manipulation will increase subjects' concerns for their social image, leading them to act more in line with the views held by society. We, hence, expect the relation between social norms and behavior to become stronger. We find that, on average, the relation between social norms and behavior becomes stronger, in line with our

conjecture. This change, however, does not come at the expense of personal norms. The relation between personal norms and behavior not only survives, but we find no evidence that it decreases at all. Together, these results show that personal norms are strong predictors of behavior across different contexts, and support them as a fundamental behavioral motive.

To further substantiate the importance of personal norms and the validity of our utility framework, we pit our utility framework against another one, in which subjects care only about social norms and their monetary payoff (see, e.g., Krupka and Weber, 2013). Our fundamental assumption is that people do not only care about social, but also personal norms. Hence, we investigate whether the addition of personal norms increases the predictive power of the estimated models. We find that adding personal norms significantly increases the predictive fit for all games, across both the PRIVATE and the SOCIAL treatment. This supports the central assumption of our framework, and shows that the inclusion of personal norms increases our understanding of economic behavior.

Finally, we test the robustness of our main findings and underscore their relevance with the following steps. First, we argue and provide evidence for why our results cannot be explained by a preference for consistency (Falk and Zimmermann, 2018). Second, we relax the assumption from our main analysis where we posit that people care about their *beliefs* about the social norm as a natural comparison to personal norm. Instead, we assign a common social norm to all subjects in line with Krupka and Weber (2013). We repeat our analyses with this approach, and show that *all* our results stay robust. Third, to complement these results, we run another experiment with a different set of subjects where we elicit the personal and social norms for: i) the same four games as in the main experiment, and ii) seven additional games as well as a battery of vignettes representing real-life economic situations. We show that the results on the two norms as well as on their relation in the four main games stay remarkably close to those in our our main sample, indicating that this is a stable finding within comparable populations. Moreover, we find substantial heterogeneity between the two norms across all the additional games and vignettes, showing that differences between the two norms are not restricted to the four games in our main experiment. In combination with our main findings, this suggests that personal norms are a relevant behavioral predictor across a wide range of economic contexts.

Our study contributes to the literature investigating the effect of social norms on economic behavior (Krupka and Weber, 2009, 2013; Kessler and Leider, 2012; Gächter et al., 2013; Banerjee, 2016; Kimbrough and Vostroknutov, 2016; Agerström et al., 2016; Krupka et al., 2017). We show that, alongside social norms, personal norms have a strong and robust relation to economic behavior. Most importantly, we show that personal norms *complement* social norms. To the best of our knowledge, the only economic study that so far considers personal norms is Burks and Krupka (2012). In one of their findings, the authors report that an overall misalignment between personal ethical opinions (a concept equivalent to personal norms) and social norms in the context of whistle-blowing is correlated to employees' job satisfaction and behavior in an "advice game". Although their study sheds some first light on the topic of personal norms, it

does not identify their relation to any particular behavior they might govern. In our study, we employ a novel design which allows us to cleanly *estimate and establish* a relation between social and personal norms on the one side and behavior on the other.

Our findings also advance the literature on image concerns, in particular in the domain of prosociality. We connect to studies on self-image (Dana et al., 2007; Gneezy et al., 2012; Grossman and Van Der Weele, 2017; Falk, 2017; Bašić et al., 2020), as personal norms hinge on inner enforcement mechanisms which rely on the image or concept one has of herself. Moreover, we contribute to the understating of social image concerns (Andreoni and Petrie, 2004; Alpizar et al., 2008; Ariely et al., 2009; Andreoni and Bernheim, 2009), as we report evidence underscoring the relevance of social norms in settings in which social image concerns are high.

Finally, our findings also relate to signaling models which capture the relation between social norms, image concerns and behavior (Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009; Bénabou and Tirole, 2011). Bénabou et al. (2018) consider a setting with multiple audiences that have conflicting norms. Our results speak to the relevance of such settings, as the presence of an internal audience (judging according to personal norms) and an external audience (judging according to social norms) is common to many economic decisions.

3.1 Social and personal norms-dependent utility framework

We start by defining the two concepts that build the cornerstones of our framework, social and personal norms. Regarding *social norms*, we closely follow the approach of Krupka and Weber (2013), and stay in line with other seminal work on the topic (see Elster, 1989; Ostrom, 2000; Bicchieri, 2005). We conceptualize them as the collective perceptions among members of a group or society, regarding the appropriateness of different actions in a given situation.[1] In this sense, they represent shared understandings about actions that are permitted or prohibited. They hinge on expectations of other and can be enforced by external sanctions (Bicchieri, 2005). Importantly, the definition we use implies that it is possible to attach a socially accepted value which indicates how appropriate an action in a given situation is according to the viewpoint of the respective group or society.

In contrast to social norms, personal norms represent one's *private* beliefs about the appropriateness of an action. To define them, we follow Schwartz (1973, 1977) who argues that personal norms come from intrinsic values and deviations therefrom are subject to intrinsic sanctioning tied to self-concept, e.g., self-depreciation (see also Schwartz and Fleishman, 1978; Cialdini et al., 1991). In this sense, personal norms do not hinge on others' expectations to follow them (Bicchieri, 2005); hence, they can depart from social norms, reflecting one's individual values.

[1]Note that this conceptualization describes injunctive social norms, i.e., prescriptions of how one *ought* to behave, and not descriptive social norms, i.e., how people usually behave. While both can influence behavior (see, e.g., Cialdini et al., 1990; Krupka and Weber, 2009; Bartke et al., 2017), the main focus in economics has been on injunctive social norms (see, e.g., Burks and Krupka, 2012; Krupka and Weber, 2013; Gächter et al., 2013; Banerjee, 2016; Kimbrough and Vostroknutov, 2016; Krupka et al., 2017). More importantly, injunctive social norms provide the ideal conceptual counterpart to personal norms.

Following this line of reasoning, we define *personal norms* as a person's individual perceptions regarding the appropriateness of different actions in a given situation, irrespective of the opinion of others. Consequently, we assume that it is possible to attach a personal value to how appropriate each action in a given situation is.

The above definitions imply that a social norm is a *commonly* held value, while personal norms can differ across people. While an individual should have more or less perfect insight in what her own personal norm is, she can only rely on her belief about the social norm which might not always coincide with the actual social norm. In fact, elicitations show rather heterogeneous beliefs about social norms, which are then commonly aggregated using some measure of the central tendency of those beliefs (see, e.g., Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016; Krupka et al., 2017). Thus, if individuals act on their belief and not on the actual social norm, using a unique value for the social norm could potentially misidentify its effect. For this reason, in our utility framework and analysis, we rely on what subjects *think* is appropriate from the viewpoint of society (perceived social norm), and what they themselves believe to be appropriate (personal norm). We find this to be a more natural way to compare personal and social norms. We later relax this assumption and repeat our entire analysis by assuming that people care about the commonly perceived (average) social norm instead of their belief about the social norm.

We now describe our utility framework. An individual takes an action a_k from a set of possible actions $A = [a_1, ..., a_k]$. She cares about: i) the monetary payoff $\pi(a_k)$ she gets from the action, ii) her belief about the appropriateness of the action from society's view $S_i(a_k)$, iii) and her own private belief about the appropriateness of the action $P_i(a_k)$.[2] $S_i(a_k)$ and $P_i(a_k)$ are functions that assign an appropriateness score in an interval $[-1, 1]$ to each action. $S_i(a_k)$ represents the perception about the commonly shared view in society and, hence, describes the subjects' belief about how socially appropriate or inappropriate it is to perform a certain action. Similarly, $P_i(a_k)$ describes the subjects' belief about how appropriate or inappropriate it is to perform an action from her own viewpoint. In both cases, a negative score means that the action is perceived as inappropriate, whereas, if the score is positive, the action is considered as appropriate. The utility function of an individual is then simply given by:

$$u_i(a_k) = V(\pi(a_k)) + \gamma S_i(a_k) + \delta P_i(a_k). \tag{3.1}$$

Here, $V()$ is the utility derived from money. The two parameters $\gamma, \delta \geq 0$ represent the tendency or concern to follow the social and personal norm, respectively. They are zero for an individual who is entirely untroubled by the two. The larger they are, the more an individual is influenced by the respective appropriateness ratings. While an individual might care about both norms, she could also be highly concerned by the social appropriateness of an action and not by the personal appropriateness, or the other way around. Importantly, we posit that the extent to

[2] We build on the framework from Krupka and Weber (2013), who assume that people care about their monetary payoff and the social norm. In our framework, we assume that people also care about a second normative principle, personal norms. A similar idea is also presented in Burks and Krupka (2012).

which people's behavior is determined by a norm is malleable. Specifically, Bicchieri (2005) argues that "situational factors may increase the effect of norms on behavior by making a norm salient"; hence, we assume that γ and δ can be affected by the environment (see also Berkowitz and Daniels, 1964; Schwartz and Fleishman, 1978; Rutkowski et al., 1983; Cialdini et al., 1991). We utilize this assumption for our manipulation of social norms as described below.

3.2 Experimental design and predictions

Our experimental design consists of two parts: an online experiment and a laboratory experiment. Each subject participated in both the online and the lab part, which were separated by a considerable time lag (approximately 4 weeks). In both parts, subjects went through four different games, which we describe in the next section. The aim of the online experiment was to elicit subjects' social and personal norms for the four games along with other variables (see Section 3.2.2). In the lab, subjects played the four games, either in a PRIVATE or a SOCIAL treatment in a between-subjects design (see Section 3.2.3). We conclude this section with our predictions for the experiment.

3.2.1 Games

We have chosen four relevant games which cover different economic settings and a wide range of motives: dictator game, dictator game with tax, ultimatum game, and third-party punishment game. The dictator game (Kahneman et al., 1986; Forsythe et al., 1994) is one of the most widely studied experimental setups and captures an individual's prosocial behavior in the absence of strategic interaction. The dictator game with tax extends this setup to a broader range of motives as it introduces a conflict between competing fairness principles. The ultimatum game (Güth et al., 1982) is a widely-used paradigm that (in contrast to the first two games) investigates fairness concerns in strategic settings. Finally, the third-party punishment game (Fehr and Fischbacher, 2004) is a more novel, but highly influential setting which studies norm-enforcement and altruistic punishment. We used role uncertainty in the games where there is a passive player: dictator game, dictator game with tax and third-party punishment game, as detailed below. By studying these four games, we intend to demonstrate that our results apply across multiple economically relevant settings.

Dictator game In the dictator game (DG), two participants are randomly matched together. Both decide how they would split an endowment of €10 (in intervals of €1), if they were assigned to the role of Dictator. This decision is private and both decide without knowing what the other would implement. Then the role assignment is disclosed: one participant is assigned to the role of Dictator and the other to that of Recipient. The decision of the actual Dictator is implemented.

Dictator game with tax The dictator game with tax (DGT) is identical to the DG above, except that both subjects now decide how they would split €12 in the role of Dictator, and any

amount sent to the Recipient is reduced by 40% (the tax). Subjects can send amounts in €1.50 increments (€1.50, €3, ..., €12). Note that sending €0 maximizes the sum of payoffs, while sending €7.5 ensures equal earnings for both players (€4.5) and sending €6 equalizes the two shares before taxes.

Ultimatum game In the ultimatum game (UG), two participants are randomly matched together and assigned the roles of Proposer and Responder. The Proposer gets €10 and offers any integer amount from €0 to €10 to the Responder. If the Responder accepts the offer, the €10 are divided as suggested by the Proposer. If, however, she rejects the offer, both participants earn nothing. We elicit the Responder's choice using the strategy method (Selten, 1965): the Responder has to state the minimum offer she wants to accept. Any offer greater or equal to the declared amount is accepted, while those below are rejected. The payoffs are determined by matching the Proposer's actual offer with the choice of the Responder. In this game, we are interested in Responders' rejection behavior.[3]

Third party punishment game In the third party punishment game (TPP), three subjects are randomly matched together. One of them is assigned to the role of Dictator. The other two subjects both have indicate how they would decide if assigned the role of Third party, before finding out whether they are actually the Third party or the passive Recipient. The Dictator gets €10 and can give either €0, €2 or €5 to the Recipient. The Third party can punish the Dictator. She gets €5 and can reduce the Dictator's payoff by €3 for each punishment point she assigns, with the Dictator's payoff being bounded below by €0. Each punishment point costs her €1. We elicit the Third party's choice using the strategy method (Selten, 1965): the Third party has to assign punishment points for each possible choice of the Dictator (€0, €2 or €5). The decisions are private and all three subjects decide without knowing what the other subjects decided. Punishment points are assigned according to the actual choice of the Dictator and the punishment choice of actual Third party. In this game, we are interested in Third-parties' behavior.

3.2.2 Online experiment

The online experiment was conducted in Qualtrics. We sent the link to the online experiment in the invitations for the laboratory experiment. The invitations were sent out four weeks before the first session of the lab experiments which took place on three consecutive days. Subjects had six days to complete the online experiment. This means that subjects completed the online experiment between 30 and 23 days before their lab session. This long time lag was specifically chosen to reduce subjects' memory regarding the online tasks and their exact answers, once they came to lab. At the beginning of the online session, participants generated a code which we

[3]While the behavior of Proposer is also interesting, the effects of norms are not straightforward to identify. In particular, the behavior depends on the Proposer's personal and social normative perceptions, as well as her beliefs about the Responder's action, which is again driven by her personal and social normative perceptions for her situation. Hence, to test our utility framework, we focus on the behavior of Responders.

used to match their data between the online and the lab session. Then, they proceeded to the main task: the elicitation of their beliefs about social and personal norms in the four games, as described below. The elicitation of norms was organized in two blocks: a block with personal norms, and a block with social norms. The order of the two blocks as well as the order of the games within each block was randomized. Each block started with an explanation of the task and an example. While solving the first block of norm elicitations, subjects were unaware of the task in the upcoming second block. For example, if they went through the personal norm elicitation first, they were not aware that afterwards they would also go through the social norm elicitation. After both blocks, we collected some demographic variables.[4]

Social norms We elicited social norms using an adapted version of the procedure by Krupka and Weber (2013). We phrased the text in a manner that allows us to directly contrast personal and social norms. Subjects had to rate how socially appropriate they believed each action to be on a 6-point Likert scale. In particular, they were faced with the following text: "For each action, evaluate according to the opinion of the society and independently form your own opinion, whether it appropriate or not to choose it. "Appropriate" behavior means the behavior that you consider most people would agree upon as being "correct" or "moral"." (see Appendix C.1 for full instructions). We re-scale the answers to an interval from -1 to 1 for the analysis. Subjects received €0.30 for each answer that matched what most other subjects had chosen.[5]

Personal norms We elicited personal norms in a symmetric procedure to social norms. We asked subjects to rate how personally appropriate they believed each action to be, irrespective of the view of others. In particular, we used the following sentence: "For each action evaluate according to your own opinion and independently from the opinion of others, whether it is appropriate or not to choose it. "Appropriate" behavior means the behavior that you personally consider to be "correct" or "moral"." Also in this case, subjects answered on a 6-point Likert scale and we re-scale their answers between -1 and 1. They were asked to answer as precisely as possible with their honest opinion. This elicitation was not incentivized, as personal norms are by definition an individual value and cannot be matched to others' personal norms (see Burks and Krupka (2012) for a similar method).

3.2.3 Laboratory experiment

The main purpose of the lab experiment was to elicit subjects' behavior in the four games. Each subject played all games, and the order of the games was randomized at the individual level. We imposed perfect stranger matching, i.e., each subject could only be matched once with another given subject across the four games. Decisions were taken simultaneously. One game was randomly selected to determine the payoff. The outcomes of the games as well as the payoff

[4]We asked for subjects' gender, age, field of study, number of siblings, favorite food and favorite movie. The last two variables were an additional safeguard to be able to distinguish subjects if they had the same code, which in fact never happened.

[5]Subject could earn up to €12 from this task.

were revealed only after all subjects went through all four games.

Subjects were randomly assigned to one of two treatments.[6] In the PRIVATE treatment, subjects made their decisions for the four games in an anonymous setting. In the SOCIAL treatment, we *exogenously* manipulated the visibility of subjects' actions in order to increase their social image concerns. To this end, subjects were informed at the very beginning of the experiment that after all participants had completed all tasks, they all would have to stand up so that everyone could see and hear everyone else. A laboratory assistant would then call up each participant one after the other. Participants would then have to say their first name and what they had chosen in each of the four games. Specifically, they would have to read verbatim a text displayed on their screen containing all information regarding all the decisions they had taken. Importantly, this approach ensured that the environment during the decision-making stage was kept constant across the two treatments, and the only difference was the information about whether their behavior would become publicly known or not (see Ariely et al. (2009) and Ewers and Zimmermann (2015) for similar manipulations).

Before the start of each game, subjects had to answer to some control questions to make sure they understood the experimental instructions correctly. Once subjects completed the main part of the experiment, they went through a short series of questionnaires (see Appendix C.1). We measured participants' situational self-awareness (Govern and Marsch, 2001) and their reputational concerns (adapted from Romano and Balliet, 2017). Subjects also completed an 11-item BIG-5 questionnaire (Rammstedt and John, 2007) and an additional short questionnaire on their memory about the online experiment and some socio-demographics.

3.2.4 Procedure

The experiment was conducted at the Cologne Laboratory for Economic Research (CLER) of the University of Cologne between October and November 2019. The online experiment was conducted using Qualtrics, while the laboratory experiment was programmed in zTree (Fischbacher, 2007). Subjects were recruited via Orsee (Greiner, 2015). Our sample consists of 250 subjects that took part both in the online and lab experiment (62% female, average age 25.8 years). In Appendix C.2, we show that there was no systematic attrition between the online and lab experiment. All subjects received a show-up fee of €8, plus their earnings from the the online experiment and their earnings from the laboratory experiment. Overall, subjects on average received a payment of €17.3. The online experiment lasted between 20 and 35 minutes, while the laboratory experiment took on average 50 minutes.

3.2.5 Predictions

We have three main predictions for the results of our experiment. First, since personal norms represent internalized values which can originate from the society (see Schwartz, 1973), we expect personal and social norms to be related. As highlighted in Section 3.1, however, they do not

[6] The random assignment was done at session level, i.e., all subjects in one session were in the same treatment.

need to be identical, as personal beliefs of appropriateness can deviate from societal ones. In fact, many economic settings contain a multitude of normative principles (e.g., equality, altruism, payoff-maximization, efficiency) that could give rise to individual differences. This heterogeneity represents a conditio sine qua non for identifying the relation between the two norms and behavior.

Hypothesis 3.1 *Appropriateness ratings of social and personal norms are correlated; however, there is non-negligible individual heterogeneity between the two.*

Second, while it is well-established that social norms and monetary payoffs influence economic behavior, we conjecture that personal norms are also a driver of behavior; thus, we expect them to play an important role in explaining subjects' actions across the four games.

Hypothesis 3.2 *Personal norms play a substantial role in explaining behavior:* $\delta > 0$ *(Equation 3.1).*

Third, as explained in Section 3.1, we also posit that the weight put on social and personal norms might differ across situations. Our treatment manipulation in the Social treatment is aimed at making *only* social norms more salient. Since social norms, in contrast to personal norms, are subject to others' expectation to follow them (Bicchieri, 2005), we conjecture that increasing the visibility of actions, i.e., social image concerns, will make subjects' more concerned about the opinion of others. If there is an expectation of following the social norm, the manipulation should raise the influence of social norms on behavior.

Hypothesis 3.3 *Social norms play a more important role in the* Social *treatment compared to the* Private *treatment:* $\gamma_{\text{Social}} > \gamma_{\text{Private}}$ *(Equation 3.1).*

The increase in observability should not affect the personal norms directly. However, if social norms become more salient, they could crowd-out the effect of personal norms, since the presence of a strong competing normative principle could "override" the effect of personal norms (see Bicchieri, 2010). Thus, when analyzing how the Social treatment affects social norms, we will also test for potential indirect effects on personal norms.

3.3 Results

Our results are structured in the following way. We first give an overview of the personal and social norms across the four games and provide evidence for their heterogeneity. Then, we move to our main results and analyze how personal and social norms are related to behavior. Here, we establish the predictive power of personal norms in the Private treatment and investigate how the weights put on personal and social norms change in the Social treatment. We then perform a series of robustness checks to validate our main results. Finally, we report data from additional experiments in which we replicate the elicitation of social and personal norms for our main games, and report the elicitation for nine additional games and ten vignettes representing real-life economic situations.

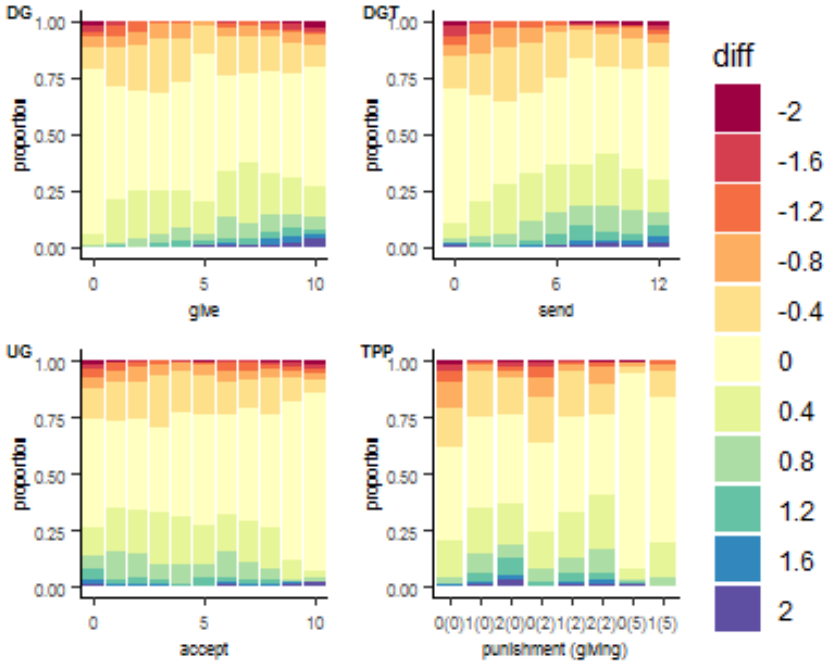


Figure 3.1: Individual difference between appropriateness ratings of social and personal norms

Note: The difference is calculated by subtracting an individual's personal appropriateness rating from her social appropriateness rating.

3.3.1 Overview and heterogeneity of personal and social norms

We start by providing evidence for the heterogeneity of personal and social norms. As argued in our predictions (see Hypothesis 3.1), we expect social and personal norms to be related but also sufficiently distinct from each other. In line with our conjecture, we find that appropriateness ratings of personal and social norms have a strong relationship. Specifically, we observe high correlations across all four games: 0.72 for the DG, 0.65 for the DGT, 0.74 for the UG and 0.71 for the TPP. However, this strong relation masks important heterogeneity. To investigate the differences at the individual level, we look at the personal and social appropriateness ratings of the available actions in each of the four games and check whether and to what extent the two ratings differ. We visualize this information in Figure 3.1. For each individual, we subtract the personal-norm appropriateness rating from the social-norm appropriateness rating for all possible actions in the four games. The difference can range from -2 to 2. A difference of 0 means that the two ratings are the same. One can easily notice that, while a difference of 0 is frequent, for a substantial amount of cases there is indeed a difference in the ratings of social and personal norms. In fact, a difference is present for 49.89% of the cases in DG, 55.64% in DGT, 49.67% in UG and 51.25% in TPP. This confirms our conjecture, and constitutes an excellent precondition to study the importance of personal and social norms for behavior.

Result 3.1 *While social norms and personal norms are correlated, there is substantial heterogeneity at the individual level across all four games.*

3.3.2 Personal norms, social norms and behavior

We now join the data of personal and social norms from the online experiment with the behavioral data from the lab. This allows us to find out whether personal norms are predictors of behavior as conjectured by Hypothesis 3.2.

To estimate our utility framework (Equation 3.1) and to capture the predictiveness of the two norm ratings, we employ a conditional (fixed-effect) logit choice model (McFadden et al., 1973). In this regression model, the dependent variable is a dummy variable indicating whether a subject chose a given action, and the independent variables are the characteristics of that action: the monetary payoff attached to the action, the individual's social appropriateness rating of that action, and her personal appropriateness rating of that action. The obtained coefficients provide estimates for the weights of our utility framework (for more details on the estimation procedure see Appendix C.2.2).

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.727***	0.338***	0.514***	0.989***	0.443***
	(0.103)	(0.051)	(0.128)	(0.158)	(0.034)
Social norm rating	0.734**	0.628**	0.561	0.628***	0.514***
	(0.365)	(0.255)	(0.358)	(0.227)	(0.130)
Personal norm rating	1.399***	0.765***	0.819**	0.712***	0.933***
	(0.323)	(0.213)	(0.338)	(0.222)	(0.124)
Observations	1,397	1,143	704	504	3,748

Table 3.1: Conditional logit estimation of choice determinants in PRIVATE treatment

Note: Estimation of conditional logit choice models with dummy variable indicating whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 3.1 provides the estimates of our model in the PRIVATE treatment. First, we look at the personal norm ratings. We find that personal norms have sizable and significant positive coefficients across all four of our games. Pooling the four games together, we observe that the personal norm coefficient remains large and significant. Turning to social appropriateness rating, we find a significant coefficient in all games except in the UG. Looking at the pooled dataset, we observe that social norms have a significant positive relation with behavior. Finally, in line with previous findings and standard economic theory, we also find that monetary payoffs are a strong and significant predictor of behavior.

Result 3.2 *Personal norms are a strong predictor of behavior across all our four games.*

In the SOCIAL treatment, we made subjects' choices observable to others in order to increase

their social image concerns.[7] According to our predictions, the manipulation should increase the importance of social norms for behavior (Hypothesis 3.3). If so, this could also have an indirect detrimental effect on the relation between personal norms and behavior.

Table 3.2 provides the estimates of a model where we test Hypothesis 3.3. We find that the coefficient of the interaction term between social norms ratings and SOCIAL is positive and highly significant in DG and DGT. Turning to UG and TPP, we do not find a significant interaction effect. Taking all games into account and pooling the dataset together, we observe that the interaction coefficient is positive and highly significant. Overall, while we observe differences across individual games, on average, we find that social norms become more important when subjects' social image concerns are increased. Turning to the interaction between personal norm rating and SOCIAL, we observe no significant effect in any of the four games as well as when pooling the dataset together. Indeed, if we look at the predictive value of personal norms in a regression that estimates their effect in the SOCIAL treatment (see Table C2 in Appendix C.2.3), the coefficients remain significant and comparable to coefficients in PRIVATE across all games. This shows that the relation between personal norms and behavior is very strong and stable.

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.763***	0.234***	0.561***	0.823***	0.358***
	(0.078)	(0.030)	(0.103)	(0.104)	(0.023)
Social norm rating	0.804**	0.313	0.585	0.527**	0.371***
	(0.343)	(0.218)	(0.358)	(0.206)	(0.120)
Personal norm rating	1.424***	0.709***	0.820**	0.750***	0.895***
	(0.323)	(0.203)	(0.339)	(0.215)	(0.119)
Social norm rating	1.259***	0.893***	0.316	-0.251	0.748***
× SOCIAL	(0.426)	(0.286)	(0.503)	(0.316)	(0.173)
Personal norm rating	0.182	-0.000	-0.176	0.372	0.091
× SOCIAL	(0.455)	(0.293)	(0.478)	(0.338)	(0.179)
Observations	2,750	2,250	1,397	990	7,387

Table 3.2: Conditional logit estimation of choice determinants interacted with SOCIAL treatment

Note: Estimation of conditional (fixed-effects) logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as well as an interaction term between personal and social norms ratings and a dummy for the SOCIAL treatment as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Result 3.3 *The relation between social norms and behavior is on average stronger in the* SOCIAL *in comparison to the* PRIVATE *treatment. The relation between personal norms and behavior remains stable.*

Next, we put our utility framework as well as the predictive value of personal norms through a

[7] As an additional step to test the validity of our manipulation, we also include a questionnaire to assess the reputation concerns, i.e., the social image concerns, of subjects (Balliet et al., 2009). We find that subjects in SOCIAL are indeed more concerned about others' opinions than subjects in PRIVATE (two-sided t-test, $p < 0.001$, $N = 250$).

further test. To do so, we pit our model against another model in which subjects care only about their monetary payoff and social norms. We test whether adding personal norms to this model leads to an improvement in predictive power. We compare the Log-likelihood measures of the estimation of these two models for each of the four games as well as for the pooled dataset. As the model with personal norms adds another predictor, we also report the Bayesian Information Criterion (BIC) which penalizes for an increase in the amount of predictors. Table C3 containing this information can be found in Appendix C.2.3. Overall, we find very strong support for our utility framework. In the PRIVATE treatment, all comparisons of Log-likelihoods are in the favor of the model with personal norms. We find a highly significant difference for each comparison, indicating that the addition of personal norms significantly increases the predictive fit of the regression models. This is also supported by the comparison of the BIC with only one exception (UG). Looking at the SOCIAL treatment, again the same finding emerges. All models with personal norms have a significantly better fit according to the Log-likelihood measure. This is again supported by the BIC, with one exception (UG).

Result 3.4 *Adding personal norms to a model which takes only monetary payoff and social norms into account significantly increases its predictive power.*

3.3.3 Robustness checks

After having established our main findings, we check their robustness in two ways. First, we rule out that the predictive value of personal norms is due to a preference to behave consistently with the answers given in the online experiment. Second, we confirm that our results do not depend on the specification we use for the concept of social norms. In particular, we re-run our analysis and, instead of using an individual's belief about the social norm, we take the average across everyone's beliefs.

Consistency. One could argue that our novel predictor, personal norms, is related to behavior due to a desire to act consistently. In particular, if people have a preference for consistency (see Falk and Zimmermann, 2018), they might want to behave in line with what they stated to be the personally most appropriate behavior in the online experiment. Our experiment was designed to minimize such concerns. During the online session, subjects answered to more than 80 items, including both personal and social norms, as well as the post-experimental questionnaire. Additionally, there was a time lag of approximately 4 weeks until the lab experiment. Hence, it is unlikely that subjects had a precise recollection of the specific answers given in the online session when making their decisions in the lab. Nevertheless, to remove any further concerns, we asked subjects at the end of the lab experiment how well they remembered the online experiment on a Likert scale from 1 (not at all) to 7 (extremely well). We test whether the predictive value of personal norms stays robust when removing those who have a good recollection of the online experiment. To this end, we re-estimate our utility framework without subjects who claimed they had a good recollection (answers 6 or 7 on the Likert scale; see Table C4 in Appendix).

Here, we pool the PRIVATE and SOCIAL treatment together for reasons of statistical power.[8] Furthermore, we also take a more extreme approach, and keep only those who reported having troubles remembering the online experiment (answers below the midpoint of the Likert scale). As this strongly reduces the sample size, we only estimate our framework using the entire dataset. Our results remain robust in all regressions.

Average social norm rating. As we argued in the utility framework, we believe that a person will act upon her belief about the norm, rather than the actual social norm (which she might fail to guess). Indeed, our dataset was conceived to obtain individual values for both the personal and social norms, allowing us to contrast the two. Here, we take a different approach, and — in line with Krupka and Weber (2013) — perform a robustness test where we assume that people care about the common social norm, which we calculate as the mean of all individual social appropriateness ratings (see also, e.g., Gächter et al., 2013; Kimbrough and Vostroknutov, 2016). We repeat all our regression analyses in Appendix C.2.4. All our results remain unaffected.

3.3.4 Further evidence on personal and social norms

In this section, we report evidence from an additional lab experiment in which we test the following two questions with a different set of subjects ($n = 160$). First, while, as we have shown above, our main results are quite robust, it would be reassuring if the patterns of the two norms, and in particular their relation would be a stable finding. To this end, we elicit social and personal norms for our four games with a different sample, and compare them to the ones we obtained in our main sample. Second, we investigate whether the heterogeneity between personal and social norms we observe in our four games also applies to other economic contexts. If so, this would give strong support for a broad applicability of our findings. Hence, we elicit the two norms for seven additional games and ten vignettes representing real-life economic situations. For more information about the procedure of these experiments see Appendix C.2.5.

Replication. First, we examine the results for our four main games. Social and personal norms appropriateness ratings display similar correlations to our main sample. Correlation coefficients for this and the main sample are 0.72 and 0.72 in DG, 0.58 and 0.65 in DGT, 0.68 and 0.74 in UG, and 0.63 and 0.71 in TPP. Also the patterns of heterogeneity, i.e., the difference at the individual level between appropriateness ratings for personal and social norms, are similar. The difference for this and the main sample is non-zero in 53.75% and 49.89% of cases in DG, 60.28 and 55.64% in DGT, 50.99% and 49.67% in UG and, 57.71% and 51.25% in TPP, respectively. Finally, we check whether the distribution of personal and social norm ratings for each action in the four games differs across the two samples. For a total of 78 tests, we find that the two distributions differ only in a single case, revealing a very consistent pattern for both normative

[8] Note that the coefficient estimating the relation between personal norms and behavior does not differ between PRIVATE and SOCIAL treatment (see Table 3.2).

perceptions.[9]

Result 3.5 *The distribution of personal and social norm ratings as well as their relation stay highly consistent and stable in a replication with a different set of subjects.*

Additional games. We elicited personal and social norms in seven additional games: Charitable giving game, Charitable giving game with entitlement, Dictator game with entitlement, Lying (die-roll) game, Ultimatum game with computer first move, Trust game and Public good game. We also elicited the two norms in ten vignettes capturing real-life situations, for example, "Your neighbour pays a painter under the table and thus pays no taxes". A full description of the games and the list of all vignettes can be found in Appendix C.2.5.

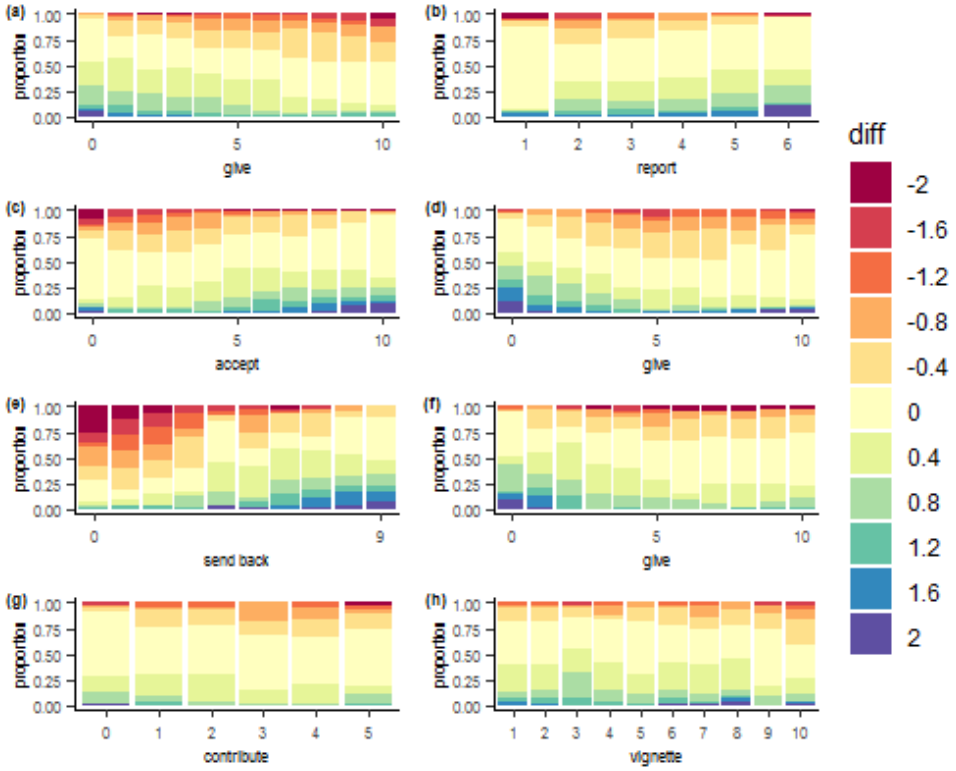


Figure 3.2: Individual difference between appropriateness ratings of social and personal norms in additional games.

Note: The following games are displayed: (a) Charitable giving game, (b) Lying game, (c) Ultimatum game with computer first move, (d) Dictator game with entitlement, (e) Trust game, (f) Charitable giving game with entitlement, (g) Public good game, and (h) 10 different vignettes. The difference is calculated by subtracting the individual personal appropriateness rating from her social norm appropriateness rating.

Figure 3.2 depicts the differences in social and personal appropriateness ratings for these additional games and vignettes. The correlation between the two ranges from 0.51 across the vignettes to 0.75 in the Public good game. The percentage of non-zero differences is again substantial, ranging from 47.87% in the Public good game to 76.74% in the Trust game. Overall, these data

[9]We run Chi-squared tests with simulated p-values over 10.000 replications and use the Bonferroni correction to account for multiple hypotheses testing at the game level for personal and social norms separately.

confirm that there is significant heterogeneity between personal and social norms. Together with our main findings, this underscores the role of personal norms as a relevant predictor of behavior across a wide range of economic contexts.

Result 3.6 *There is substantial individual heterogeneity between personal and social norms across seven additional games and ten vignettes describing real-life economic situations.*

3.4 Discussion and Conclusion

In this study, we conjecture that people, in addition to caring about social norms and their monetary payoff, also care about personal norms. We propose a simple utility framework that captures these relations and design a novel two-part experiment to estimate it. Moreover, we test the conjecture that social norms gain importance when people's actions are under the scrutiny of others, and, conversely, investigate how this affects personal norms.

We start by establishing that personal and social norms are related, but that there is substantial heterogeneity between the two at the individual level. We then estimate our framework and show robust evidence that personal norms — while taking social norms and monetary payoff into account — are strong predictors of economic behavior across four different economic games, both in a PRIVATE treatment where decisions are anonymous and in a SOCIAL treatment where social image concerns are made salient. In line with our predictions, the increase in social image concerns strengthens the relation between social norms and behavior; however, this does not come at the expense of personal norms. We then show that adding personal norms to a standard social norms framework — where people care only about social norms and their monetary payoff (see, e.g., Krupka and Weber, 2013) — significantly increases the predictive power of the estimated framework. Finally, we successfully replicate our findings regarding the relation and the heterogeneity between the two norms for our four main games, and show that this heterogeneity exists also across seven additional games and a battery of real-life economic situations.

The evidence we present in this study clearly shows that personal norms are *strong predictors* of behavior in economic settings, and moreover, it supports them as a key motive of economic decision making. Since we observe that personal norms are distinct from social norms across a large array of games and vignettes, the implications of these findings are likely to extend to a wide array of economic contexts. Taken as a whole, our results imply that future research should consider also personal norms when investigating normative precepts and their effect on economic behavior.

While our findings highlight the relevance of personal norms, it is important to stress that we do not belittle the role played by social norms. On the contrary, our results take both norms into account and provide insights on how the two norms interact and how they relate to behavior. In line with the existing literature, we find that social norms play an important role; however, the fact that the addition of personal norms increases the predictive power of the estimated models

indicates that personal norms are complementary to social norms in predicting behavior. Apart from offering support to our utility framework, this implies that by ignoring personal norms and focusing only on social norms, we are worse off in forecasting how people will behave in economic settings. These findings can have important implications, for instance, in the design of behavioral interventions. If people's behavior is co-determined by private and social norms, an intervention targeting only social norms might lack effectiveness or even fail completely. In such a situation, considering and understanding personal norms is decisive to design a successful intervention.

Finally, apart from offering evidence that both types of norms influence behavior, we also shed light on how they interact and how the focus can be shifted from one to the other. Our findings from the SOCIAL treatment indicate that increasing social image concerns enhances the importance of social norms for behavior. This supports our conjecture that situational factors can make a particular norm salient (see Bicchieri, 2005; Berkowitz and Daniels, 1964; Schwartz and Fleishman, 1978; Rutkowski et al., 1983; Cialdini et al., 1991). While we cannot dismiss the possibility that stronger manipulations might decrease the influence of personal norms, this finding suggests that personal norms are rather robust (see Bicchieri, 2010), and "overriding" this motive is far from trivial.

Chapter 4

The Dark Side of Experts: Ethical Decision-making under Asymmetric Information in Teams

Firms often face the trade-off between higher profits and potential negative externalities. For example, installing a cheap particulate filter on a diesel car will increase profits. However, it might also harm consumers' health or the environment at large. These decisions are typically taken by teams of decision-makers from boards of directors to project teams and groups of managers lower down the hierarchy. There are some good reasons for this: we think that teams will come to better decisions than individuals by aggregating different skills and sources of knowledge. However, this also implies that information about the presence or the extent of a negative externality will be distributed *asymmetrically* within the team. Some decision-makers in the teams will have more precise information (I will call them "*experts*") and others will only posses less precise information (the "*non-experts*"). In the example, the technical management or the engineers will have a better idea about the health or environmental risks of the cheaper filter, while the sales or marketing management will have less precise information.

Economists have been studying informational asymmetries for a long time. Indeed, asymmetric information is at the basis of modern contract theory (Bolton et al., 2005) and has been identified as one of the major reasons for why countries go to war (Jackson and Morelli, 2011). In markets, they can lead to big welfare losses and even to their complete breakdown (Akerlof, 1978). Organizations have to deal with informational asymmetries on a daily basis in designing incentives for their employees and in coordinating their efforts towards a common goal (see Bergh et al., 2019, for a review). Importantly, even when teams work towards a joint objective, these asymmetries are not always resolved and can lead to sub-optimal decisions (Stasser and Titus, 1985; Brodbeck et al., 2007).[1]

[1]In psychology and management, a vast literature on the "hidden profile" paradigm shows that groups are often not able to aggregate all information at the disposal of single group members due to several biases occurring

This chapter takes asymmetric information about negative externalities as a starting point and investigates its consequences for ethical decision-making in teams.[2] I use a laboratory experiment to isolate three channels that might lead to more unethical, i.e., socially harmful, behavior in these settings and ask. Will the decision-makers with better information about a potential negative externality instigate more unethical behavior? Will they intervene when their private information implies that the risk of the negative externality is high?

A major challenge when studying settings in which the information available to economic actors varies is that their incentives vary with it. A laboratory experiment ensures not only tight control over the incentive structure that individuals face, but, more importantly for the purpose of this study, it offers the unique opportunity to manipulate the information individuals receive in a clean way. I capitalize on these features and design an experiment that captures the core of the trade-off described above. Two decision-makers with perfectly aligned monetary incentives form a team and can choose between two options. One option is more profitable, but might have a negative externality on a third party (installing the cheap filter in the example). The other option is less profitable, but ensures that no externality is generated (installing a more expansive, safer filter). There is uncertainty about the presence of the negative externality, which occurs in only one of two - ex ante equally likely - states of the world. Both decision-makers receive noisy signals about the state of the world they are in. They then have to agree on which option to implement following a structured decision protocol. One of the two decision-makers is randomly picked to start, she can propose to implement one of the two options or delegate the decision to the other decision-maker. If she delegates, the other decision-maker takes a definitive choice between the two options. Otherwise, the other decision-maker receives the proposal and can either agree or make a counter-proposal. The process goes on until the two agree on which option to implement.

In the main treatment (AsymmInfo), one of the decision-makers obtains a more precise signal, which makes her the *expert*.[3] I compare the decisions of experts in AsymmInfo with those of decision-makers in a HighInfo condition, in which both decision-makers receive the more precise signal, and the decisions of non-experts in AsymmInfo with those of decision-makers in a LowInfo condition, in which both receive the less precise one. This allows me to compare the decisions taken by decision-makers with the same Bayesian posterior about the presence of the externality, while varying only whether information is distributed symmetrically or asymmetri-

at both the group and individual level (Stasser and Titus, 1985, 2003; Wittenbaum et al., 2004; Sohrab et al., 2015). Other reasons for why asymmetries might not be ironed out come from temporal or financial constraints that often dictate when and how decisions are taken.

[2]Firms might not always have the right incentives to reduce informational asymmetries in these situations. Indeed, this seems to have played an important role in the "Dieselgate" scandal in which Volkswagen's executives claimed to have been unaware of the misconduct concerning the manipulation of the emission-control software during laboratory emission tests (BBC, 2015) and attributed the blame to engineers working on the software (O'Kane, 2015).

[3]Experts in my experiment are decision-makers who can form a more precise Bayesian posterior about the presence of the externality. While this lies at the core of the notion of expertise, there can be several reasons why an individual might reach a more precise Bayesian posterior. One might derive it from past experience, or have better access to information, or a superior technology to elaborate it. By endowing a decision-maker with more precise information, I choose a clear-cut way to operationalize and manipulate expertise.

cally.

By implementing a structured protocol for the decision process, I can isolate and study three potential behavioral mechanisms that might lead to more unethical choices under asymmetric information. The first mechanism concerns the cases in which the non-expert delegates the decision to the expert. Since delegation has been found to be an effective way to avoid the responsibility of selfish actions (see, e.g., Bartling and Fischbacher, 2011), the non-expert might delegate the decision to the expert to free herself from this burden. If the expert interprets delegation as a sign of neglect towards the well-being of the third party, she might exploit this leeway to behave more unethically. On the other hand, if the expert feels responsibilized, she might behave more ethically. The second mechanism concerns the cases in which the expert makes the first proposal. Since the non-expert does not learn the expert's private signal, the social image concerns of the expert vis a vis the non-expert are lower compared to the symmetric case in which both decision-makers have the same information. Hence, the expert might feel free to behave more unethically and make more unethical proposals compared to the symmetric case. Finally, when the expert receives a proposal from the non-expert, she might ignore her private information and not intervene to prevent unethical outcomes from happening. In contrast, if the information was shared, she would not do so.

I find that experts do not exploit the leeway given by delegation to take more unethical choices: if anything, they behave more ethically. Additional data suggest that this might be driven by the greater sense of responsibility they feel because of their more precise information. Experts do not initiate unethical behavior via their proposals either. On the contrary, they seem to condition their initial proposal on the signal they share with the non-expert, ignoring their private information. As I discuss in Section 4.4, this suggests that experts base their social image concerns on the shared information in line with the literature on hidden profiles (Stasser and Titus, 2003; Wittenbaum et al., 2004). However, this also also leads experts to ignore their private information and agree to implement unethical choices when the risk of a negative externality is high. The fact that experts do not actively cause more unethical choices, but fail to prevent them by passively agreeing to implement them, is coherent with an omission-commission bias (Spranca et al., 1991), which I further elaborate on in Section 4.4. Overall, high negative externalities are generated despite the presence of the expert. In fact, providing additional information to one decision-maker alone does not reduce negative externalities, while harmful choices are greatly reduced if both decision-makers receive more precise information.

Ethical decision-making in teams under asymmetric information has two sides. On the bright side, experts do not actively initiate more unethical behavior. Having additional information raises their feeling of responsibility and can lead them to make more ethical decisions. On the dark side, however, they do not step in to prevent unethical behavior and ignore their additional information, even if it points to high potential risks. These results show that asymmetric information can have adverse effects on ethical decision-making in teams by providing "wiggle room" and excuses to behave unethically. The implications for the design of information structures and

decision processes are serious. Organizations and firms should prevent experts from adopting a passive role in the decision process. Experts should be nudged to take the lead in the decision process and speak up first to avoid the generation of negative externalities.

4.1 Related Literature

A well-established finding in the papers on team decision-making is that teams behave more rationally and more selfishly compared to individuals (Charness and Sutter, 2012). Teams are more selfish in dictator games (Luhan et al., 2009), ultimatum games (Bornstein and Yaniv, 1998), trust games (Kugler et al., 2007; Cox, 2002), and gift-exchange games (Sutter and Kocher, 2007). Groups have also been found to behave less morally in other contexts. People lie more in collaborative settings (Soraperra et al., 2017; Weisel and Shalvi, 2015), with team incentives (Wiltermuth, 2011; Conrads et al., 2013), and in groups (Sutter, 2009; Kocher et al., 2017; Barr and Michailidou, 2017; Chytilova and Korbel, 2014). However, little is known about how heterogeneity in teams affects such outcomes. This chapter extends the rich literature on team decision-making by introducing asymmetry in the information structure.

In a related paper, Ellman and Pezanis-Christou (2010) study how different organization and communication structures affect unethical choices. They find that vertical, hierarchical structures are more conducive to unethical outcomes than horizontal structures in which consensus is required. Similarly, Falk and Szech (2013) study the role of pivotality within organizational structures and find that the possibility of diffusing responsibility favors immoral outcomes. I extend on this research by investigating whether the way information is distributed in a setting with no explicit hierarchies might lead to more unethical outcomes. In particular, I operationalize the notion of expertise by endowing one of the two decision-makers with more precise information. This enables me to capture a major component of what it means to be an expert in a clear and simple way. Moreover, it allows to construct a straightforward experimental design. A similar notion of expertise has been used in research about abstention and delegation in voting (see Morton and Tyran, 2011).

Uncertainty about the consequences of one's decisions for others has been found to be detrimental to moral or ethical behavior. Exley (2015) shows how risk is used as an excuse not to donate to charitable giving. People also self-servingly interpret ambiguity to justify unfair behavior (Haisley and Weber, 2010). In general, moral grey areas are often exploited to one's own advantage (Dana et al., 2007) and fairness arguments are invoked to favor one's self-interest (Konow, 2000). However, most of these papers focus on individual behavior or settings in which there is no real interaction between individuals. I build on this literature and focus on informational asymmetries that, as I argue above, are the norm in organizations.

Last but not least, this chapter contributes to the field of organizational behavior and, in particular, to that of behavioral and business ethics (Dana et al., 2012; Treviño et al., 2014) by pointing to a source of unethical behavior which organizations should pay attention to.

4.2 Experimental Design

The experimental design consists of a main treatment with asymmetric information (AsymmInfo) and two baseline treatments (LowInfo and HighInfo). The two baseline treatments allow for a clean comparison of the behavior of *experts* and *non-experts* in the main treatment. In LowInfo, decision-makers take decisions under the same Bayesian posterior as non-experts in AsymmInfo, while in HighInfo they have the same Bayesian posterior as experts. This means that the only difference in these comparisons is whether the decision is taken under symmetric or asymmetric information in the team. In the following, I describe the basic structure of the game.[4] Next, I explain the structure of the signals decision-makers receive and how this differs across treatments. Then, I describe which additional experimental measures were elicited before and after the lab experiment. Finally, I describe the three mechanisms through which asymmetric information might lead to more unethical behavior and formulate my main hypotheses.

Game. In the main part of the experiment, three subjects are randomly grouped together. Two subjects are randomly assigned the role of decision-makers (**X** and **Y**). They build the team taking the decision. The remaining subject is a passive **third party**. She is the subject who potential suffers an externality. X and Y have to decide jointly which of two options they want to implement. Their monetary incentives are perfectly aligned. One option yields a higher payoff for them, but it potentially reduces the payoff of the third party, while the other option yields a lower payoff for X and Y, but ensures that the third party is not harmed. Table 4.1 depicts the payoffs of all three players associated to the two options. The payoffs depend on the state of the world subjects are in. There are two ex-ante equally likely states of the world: A and B. *Option 2* yields a sure payoff of 5 for all three players in both states of the world. *Option 1* yields a sure payoff of 6 for X and Y. In this case, the third party earns 5 in State A, but 0 in State B. In other words, Option 1 can have an externality on the third party whose payoff in State B is reduced from 5 to 0.

	State A	**State B**
Option 1	$6, 6, 0$	$6, 6, 5$
Option 2	$5, 5, 5$	$5, 5, 5$

Table 4.1: Game

Note: The payoffs are displayed in the following order: $\pi(X), \pi(Y), \pi(thirdparty)$

X and Y receive a noisy signal about the state of the world they are in, as detailed below. Then, they have to agree on which of the two options to implement based on a structured decision protocol. One of the two decision-makers is randomly selected to be the first mover. She can propose to implement Option 1 or Option 2, or delegate the decision to the other decision-maker. If she delegates, the other decision-maker takes a definitive choice between Option 1 and Option 2. Otherwise, her proposal is forwarded to the other decision-maker (the second mover), who

[4] Instructions translated from German can be found in Appendix D.1.

Figure 4.1: Urns

can decide to either agree with the proposal or make a counter-proposal suggesting the other option.[5] This procedure goes on until the decision-makers agree on one of the two options. In case no agreement is reached within five minutes, all subjects in the the group (X, Y, and third party) get a payoff of 0.[6] This protocol is designed to analyze and cleanly isolate the behavioral mechanisms which might lead to more unethical choices by comparing the behavior of first and second movers across treatments, as detailed below. Importantly, the two decision-makers X and Y are told where the other decision-maker is seated in the lab (i.e., in which cabin). This is done to heighten their social image concerns, as this is crucial to the mechanisms described below.

Signal. State A and State B have the same ex-ante probability ($50\% - 50\%$). Both decision-makers receive a noisy signal about the actual state of the world. They draw balls without replacement from the urn corresponding to the state they are in. The urn of State A contains 3 red balls and 2 blue balls. The urn of State B contains 2 red balls and 3 blue balls (see Fig. 4.1).

Treatments. The main treatment of interest is the one in which one decision-maker has more information than the other, i.e., she is the expert. In this treatment (AsymmInfo), both decision-makers draw *one* common ball from the urn corresponding to the state of the world they are in. Then one of them (the expert) draws an *additional* ball without replacement, i.e., she receives a more precise signal. In the LowInfo and the HighInfo treatments, decision-makers draw either *one* or *two* common balls (without replacement), respectively. The choices of decision-makers in LowInfo are compared with those of non-experts in AsymmInfo. Those of decision-makers in HighInfo with the ones of experts in AsymmInfo. Hence, decision-makers in any given comparison can always derive the same Bayesian posterior from their signals. What changes is whether information is distributed symmetrically or asymmetricaly in the team.

Additional measures. After decision-makers have taken a decision, they go through a questionnaire. First, I check whether they are able to infer what the correct Bayesian posterior is for given signals (see Appendix D.2 for the analysis of these data). After that, I ask them to report on a 5-point Likert scale how responsible they feel for the decision taken in their group. Then, they answer similar questions regarding their feeling towards the decision taken. The feelings elicited are: guilt, shame, excitement, contentment, happiness, amusement, and pride.[7] Finally,

[5] The second mover is not given the option of delegating to the first mover, since this would trivially correspond to agreeing with her proposal.

[6] No group ever reached the threshold of five minutes. In fact, in most cases proposals of first movers were accepted immediately.

[7] The latter three are used as filler items, as is usually done to avoid demand effects.

I elicit basic demographic variables: age, gender, along with field and semester of study.

Hypotheses. The experimental design just laid down allows me to analyze three behavioral mechanisms that might lead to more unethical outcomes with asymmetric information. The first is delegation, for which there are two competing hypotheses. Notice that, if a decision-maker wants to ensure that the ethical option (Option 2) is implemented in this experiment, she should never delegate the decision to the other decision-maker. Likewise, if she only cares about her own payoff, she should not exit the decision process. However, an effective way to achieve a higher payoff without bearing the responsibility of its unethical consequences is to delegate the decision to someone else (Hamman et al., 2010; Bartling and Fischbacher, 2011; Erat, 2013). In AsymmInfo, the fact that the expert has more precise information can offer a valid justification to do so. If, indeed, non-experts delegate the decision to experts two competing forces are at play. If an expert interprets delegation as a sign of neglect towards the well-being of the third party, she might choose more unethically. In this sense, delegation could offer the expert the moral wiggle-room (Dana et al., 2007) to act unethically. However, the expert might also feel a higher sense of responsibility (which is measured after the decision), because now the decision depends solely on her. This, in turn, might lead to more ethical behavior. The fact that the first mover in the decision process is randomly determined gives a clean comparison between the decision to delegate of decision-makers in LowInfo with that of non-experts in AsymmInfo.[8] Hence, in cases of delegation there are two competing hypotheses.

Hypothesis 4.1a *Experts will choose more unethically, if they interpret delegation as a sign of neglect towards the well-being of the third party.*

Hypothesis 4.1b *Experts will choose more ethically, if they feel responsibilized by their role.*[9]

Another potential source of more unethical behavior in AsymmInfo is the fact that experts are less exposed to social image concerns, given their private information. People care about their social image (Andreoni and Bernheim, 2009), i.e., about the inference others make on their level of prosociality (Bénabou and Tirole, 2006). This inference is based on the potential consequences of one's action. Since experts have additional information on the presence of the negative externality, the inference of non-experts is imprecise. Non-experts do not know on which grounds experts base their decision. Hence, social image concerns of experts in AsymmInfo are weaker than those of decision-makers in HighInfo, leading to more unethical proposals by experts. For this hypothesis, the fact that the first move is allocated randomly allows for a clean comparison between proposals of experts in AsymmInfo and those of decision-makers with the same posterior belief in HighInfo.

Hypothesis 4.2 *Experts in* AsymmInfo *make more unethical proposals compared to decision-makers in* HighInfo *under the same posterior probability, because their social image concerns are weaker.*

[8] This also reduces demand effects on non-experts to delegate the decision to experts, since they are not the only ones who can delegate ex ante.

[9] For both comparisons, decisions under the same posterior probability of the externality are considered.

Last, decisions in AsymmInfo might be more unethical, because experts fail to intervene when their private information points to a higher risk of the externality. If a non-expert proposes the unethical option, the expert might ignore a private signal pointing to the presence of an externality. In contrast, if both decision-makers were aware of such signal, this would not happen.

Hypothesis 4.3 *Experts in* AsymmInfo *ignore their private signal and agree to the unethical option, leading to more unethical choices compared to* HighInfo *under the same posterior probability.*

Note that the only implicit hypothesis made about the behavior of non-experts is that they will delegate the decision to expert, at least in some cases. I will briefly compare delegation decisions in Section 4.3, a further analysis of non-experts' behavior can be found in Appendix D.2, as this turns out not to be crucial for the analysis of unethical behavior.

Procedure. The laboratory experiment was run at the Cologne Laboratory for Economic Research (CLER) in June and July 2019. It was programmed in z-Tree (Fischbacher, 2007). Subjects were recruited using ORSEE (Greiner, 2015). The experiment lasted 40 minutes on average. One week before the lab experiment took place, subjects received the link to an online study, which they had to complete within three days. In the online study, subjects' Social Value Orientation was elicited with the SVO slider measure by Murphy et al. (2011) together with additional psychological measures.[10] The online experiment was conducted using Qualtrics and took 10 to 15 minutes. Subjects who did not complete the online study could not take part in the laboratory experiment and were excluded from any payment. A total of 357 subjects took part in the experiment (62% female, average age 25.9), 72 in LowInfo, 102 in HighInfo, and 183 in AsymmInfo.[11] Subjects were paid 14.32 € on average.

4.3 Results

I analyze the results of the experiment, closely following the hypotheses laid down in the previous section. First, I look at the cases in which the decision was delegated to experts. Then, I examine whether experts made more unethical proposals. After that, I check whether experts ignore their private information when agreeing to non-experts' proposals. I conclude by assessing the consequences of asymmetric information for the generation of negative externalities. Throughout this section, I will contrast decisions taken under the same posterior probability about the presence of the externality. In doing so, I will always refer to the posterior probability of State A ($p(bad)$), i.e., the state in which the negative externality is present (bad state). Hence, a higher probability indicates that the negative externality is more likely to be present. I will refer to Option 1 as the unethical option and to Option 2 as the ethical option. P-values for proportion

[10] Big 5 questionnaire (Rammstedt and John, 2007), Context Dependence and Independence questionnaire (Gollwitzer et al., 2006), a reduced form of the Moral Disengagement questionnaire (Bandura et al., 1996), and a modified version of the Moral Identity Scale (Aquino and Reed, 2002).

[11] 6 subjects could not finish the experiment due to technical issues and one subject had not participated in the online experiment. These subjects were excluded from the analysis.

tests are reported, unless otherwise stated.

4.3.1 Delegation

The first step to study delegation is to establish whether non-experts indeed delegate the decision to experts. Hence, I compare first-mover decisions of non-experts in AsymmInfo to those of decision-makers in LowInfo taken under the same posterior. As non-experts in AsymmInfo and decision-makers in LowInfo draw only one ball, $p(bad)$ can be either 0.6 or 0.4, depending on whether a red or a blue ball is drawn. decision-makers in LowInfo never delegate the decision right under $p(bad) = 0.6$, whereas they do so in 6% of cases (one case) with $p(bad) = 0.4$. In both cases, the frequency of delegation is not significantly different from zero (this is trivially true for $p(bad) = 0.6$, while for $p(bad) = 0.4$, the 95% asymptotic binomial confidence interval contains the zero $[-0.06, 0.19]$). Non-experts in AsymmInfo delegate 25% of the times with $p(bad) = 0.6$ and 15% of the times with $p(bad) = 0.4$. Both frequencies are statistically different from zero (95% exact binomial confidence intervals are $[0.055, 0.572]$ and $[0.032, 0.379]$, respectively). Pooling data across posteriors within the two treatment conditions, the resulting share of delegation decisions is 4% in LowInfo and 19% in AsymmInfo. Delegation is not significantly different from zero in LowInfo (95% asymptotic binomial confidence interval $[-0.027, 0.084]$), whereas it is for AsymmInfo (95% exact binomial confidence interval [0.0720.364]).[12] Overall, delegation from non-experts to experts indeed takes place in AsymmInfo, but how do experts decide in these cases?

Experts choose the unethical option 17% of the time when they are delegated to decide. If they are not, the unethical option is chosen 46% of the time in case they are second movers in the decision process and 33% of the times overall in AsymmInfo. However, as argued above, a proper comparison can only be performed by juxtaposing decisions taken under the same posterior. There are two possible benchmarks to perform this comparison. The first is to take cases where no delegation took place in AsymmInfo. The second is to use the choices of decision-makers in HighInfo.

In the first comparison, choices taken after delegation are less unethical then when there is no delegation for $p(bad) = 0.25$ ($p = 0.09$). This difference is not significant for $p(bad) = 0.5$ ($p = 0.92$). While there are no cases in which experts were delegated to take a decision with a posterior with $p(bad) = 0.75$ in AsymmInfo. Hence, experts who are delegated to take a decision are, if anything, more ethical according to this comparison. In the second comparison, there is no significant difference for $p(bad) = 0.25$ ($p = 0.6$) and $p(bad) = 0.5$ ($p = 0.19$), while, as above, the comparison for $p(bad) = 0.75$ cannot be performed. Hence, choices taken when experts are delegated to choose are not more unethical than when there is no delegation in HighInfo.[13] Overall, this lends partial support in favour of Hypothesis 4.1b. Additional data (see Appendix

[12] A comparison between LowInfo and AsymmInfo reveals no significant difference for $p(bad) = 0.6$ and $p(bad) = 0.4$ ($p = 0.13$ and $p = 0.41$), and only a marginally significant difference for the pooled observations ($p = 0.1$). There are no cases of delegation in HighInfo.

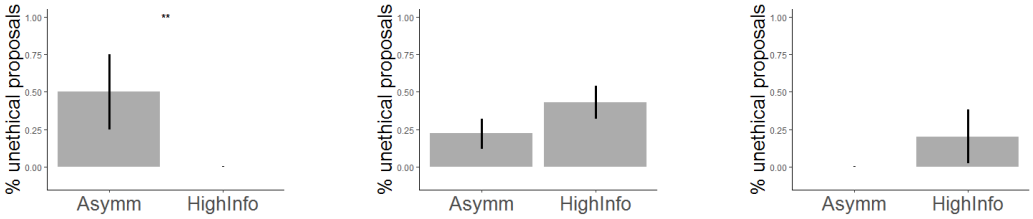
[13] Note that choices taken in the HighInfo treatment are more ethical on average, as reported below.

D.2) show that experts always feel more responsible for the choice taken in the team. They also feel guiltier and more ashamed about the choice.

Result 4.1 *When the decision is delegated to experts in* AsymmInfo*, they become, if anything, more ethical.*

4.3.2 Proposals of experts

As a next step in the analysis, I look at whether proposals of experts, i.e., their decisions as first movers, are more unethical in AsymmInfo compared to those of decision-makers in HighInfo, as suggested by Hypothesis 4.2. Also here, I compare decisions taken under the same posterior probability. Experts in AsymmInfo and decision-makers in HighInfo draw two consecutive balls without replacement from the urn. Hence, $p(bad)$ can be either 0.25, 0.5 or 0.75, if two blue balls, one blue and one red ball, or two red balls are drawn.



(a) p(bad)=0.25 ● ● **(b) p(bad)=0.5** ● ● **(c) p(bad)=0.75** ● ●

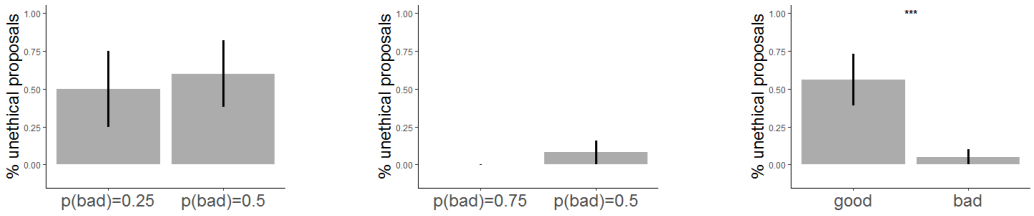
Figure 4.2: % of unethical proposals

Note: Percentage of unethical proposals under $p(bad) = 0.25$, $p(bad) = 0.5$ and $p(bad) = 0.75$ of experts in AsymmInfo and decision-makers in HighInfo (with standard errors, *** p<0.01, ** p<0.05, * p<0.1).

Fig. 4.2 depicts the share of unethical proposals in AsymmInfo and HighInfo for each of the three possible posteriors. Starting with the lowest posterior of $p(bad) = 0.25$ (a), the share of unethical proposals is 50% in HighInfo and 0% in AsymmInfo ($p = 0.028$). With a higher $p(bad) = 0.5$ (b) unethical proposals are 43% and 22% ($p = 0.173$). Finally, with the highest posterior of $p(bad) = 0.75$ (c), there are no unethical proposals of experts in AsymmInfo, while decision-makers in HighInfo propose the unethical option 20% of the times ($p = 0.251$). Contrary to Hypothesis 4.2, proposals of experts are not systematically more unethical than those of decision-makers in HighInfo. Proposals are more unethical only with a low $p(bad)$.

However, differently than in HighInfo, experts in AsymmInfo are paired with a non-expert who sees only one ball. Hence, their behavior might differ depending on whether this common signal is good or bad. Figure 4.3 compares proposals depending on the common signal. When the common signal is good (a), the expert can receive either another good or bad signal, leading to $p(bad) = 0.25$ or $p(bad) = 0.5$. Unethical proposals are 50% and 60% ($p = 0.764$). Similarly, when the common signal is bad (b), the expert can receive a good signal or another bad signal, leading to $p(bad) = 0.5$ or $p(bad) = 0.75$. In this case, unethical proposals are 7% and 0% ($p = 0.485$). Hence, experts seem to condition their proposal on the common signal they share with the non-experts. In fact, pooling proposals for the cases in which the common signal is

good or bad (c) and comparing the share of unethical proposals yields a very large difference (56% vs. 5%, $p = 0.002$).



(a) Good common signal ● **(b) Bad common signal** ● **(c) Pooled comparison**

Figure 4.3: % of unethical proposals

Note: Share of unethical proposals with a good and bad common signal, and pooling together proposals with bad and good common signals of experts in AsymmInfo (with standard errors, *** p<0.01, ** p<0.05, * p<0.1).

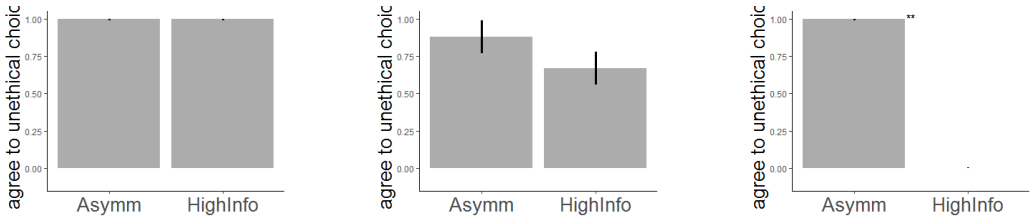
Result 4.2 *Proposals of experts in* AsymmInfo *are not systematically more unethical than those of decision-makers in* HighInfo*. Experts in* AsymmInfo *condition their proposal on the common signal. They make more unethical proposals with a good common signal and less with a bad common signal.*

4.3.3 Agreement

After having checked whether experts in AsymmInfo initiate more unethical behavior when delegated to take the decision or through their proposals, I look at what happens when they receive an unethical proposal from the non-expert. According to Hypothesis 4.3, experts might ignore their private signal and agree to implement the unethical option. This might lead to more unethical choices in AsymmInfo compared to HighInfo.

Figure 4.4 displays the share of unethical proposals which experts in AsymmInfo and decision-makers in HighInfo agree to. For $p(bad) = 0.25$ (a) agreement is at 100% in both cases, i.e., when an unethical proposal comes in, subjects always agree. Agreement for $p(bad) = 0.5$ (b) is at 88% for experts and 67% for decision-makers in HighInfo ($p = 0.269$). Finally, when $p(bad) = 0.75$ experts always agree to the unethical option, while decision-makers in HighInfo never do so ($p(bad) = 0.014$). This means that, when the probability of being in the bad state is high, unethical offers in HighInfo are never implemented and experts never propose the unethical option (see previous Section), but they always agree to the unethical option, if proposed by the non-expert. Overall, agreement to unethical proposals stands at 93.75% for experts in AsymmInfo and only at 60% for decision-makers in HighInfo ($p = 0.003$).

Result 4.3 *Experts in* AsymmInfo *agree to implement more unethical decisions than decision-makers in* HighInfo*.*



(a) p(bad)=0.25 ● ● **(b) p(bad)=0.5** ● ● **(c) p(bad)=0.75** ● ●

Figure 4.4: Agreement to unethical proposals

Note: Share of agreement to unethical proposals under $p(bad) = 0.75$, $p(bad) = 0.5$ and $p(bad) = 0.25$ of experts in AsymmInfo and decision-makers in HighInfo (with standard errors, *** p<0.01, ** p<0.05, * p<0.1).

4.3.4 Externalities

In the previous sections, I analyzed the influence of asymmetric information on unethical behavior by looking at isolated mechanisms. The next natural step is to piece these mechanisms together to look at their aggregate consequences. The central question of this chapter is whether asymmetric information leads to more unethical outcomes in team decision-making. To answer this question I now look at the ex-ante expected externality inflicted on the third party in the three different treatments. To get at this value I first calculate the size of the expected externality generated under each posterior in the three treatments, which depends on the posterior probability of the externality itself and the share of unethical choices taken under that posterior in a given treatment. Then, I multiply the values obtained by the ex-ante probability to end up with that given posterior. This allows me to compare the externalities generated under the three different information structures of my treatments. The ex-ante expected externality in AsymmInfo (0.44 €) is almost exactly the same as that in LowInfo (0.46 €). Hence, endowing only one decision-maker with more precise information does not reduce the generation of negative externalities. However, if both decision-makers have the same, more precise information as in HighInfo, the expected externality shrinks by as much as 33% (0.31€).

Result 4.4 *The presence of an expert in* AsymmInfo *does not reduce the expected externality compared to* LowInfo*, but leads to a much higher expected externality than in* HighInfo*.*

4.4 Discussion and Conclusion

This experiment explores the consequences of asymmetric information about a negative externality for unethical behavior in team decision-making. On the bright side, experts do not exploit being delegated to take the decision as an opportunity to make more unethical choices. As Result 4.1 shows, they become, if anything, more ethical when delegated to take a decision. Experts also do not initiate unethical behavior themselves through proposals (Result 4.2). Correlational evidence indicates that this might be due to an enhanced feeling of responsibility, combined with more pronounced feelings of guilt and shame, or the anticipation thereof. In fact, experts

condition their decision on the common signal they share with non-experts and seem to ignore their private information. This suggests that they think they will be judged upon what is common knowledge, i.e., that their social image concerns seem to be predominately based on the common information. Such failure to incorporate private information in team decision-making is in line with a large literature on the hidden profile problem in psychology (Stasser and Titus, 1985, 2003; Wittenbaum et al., 2004; Sohrab et al., 2015). Recent work in economics shows that people, indeed, implement actions they know to be socially harmful to appear prosocial in the eyes of others (Soraperra et al., 2019).

Experts also ignore their private information and agree to unethical behavior even when they know the probability of the externality to be high (Result 4.3). This dark side of experts leads to the generation of high externalities. This is coherent with an omission-commission bias in judgement (Spranca et al., 1991): holding consequences constant, acts of omission are considered to be less immoral than acts of commission. Similarly, in this setting, experts agree to implement the unethical option, but would not actively propose it.

Overall, giving more precise information to only one decision-maker does not reduce unethical outcomes, compared to a situation in which both decision-makers have less precise, symmetric information (Result 4.4). Instead, when both decision-makers have more precise, symmetric information, socially harmful behavior is greatly reduced (Result 4.4). The overall pattern of results highlights the perils of asymmetric information about negative externalities for ethical behavior. Asymmetric information seems to create a moral wiggle room (Dana et al., 2007) and provides justifications that offer a way to morally disengage from the situation (Moore et al., 2012). This helps people to maintain a positive self- and social image while acting unethically in their own interest (Mazar et al., 2008).[14]

These results have important consequences for the design of decision processes. Organizations that wish to promote ethical behavior should try to avoid situations in which better-informed individuals might disengage from the decision process, since this can lead them to ignore their private information. This means that experts in teams should be given priority to express their opinion and that they should be prevented from taking on a passive role.

[14]In line with this interpretation I find that decision-makers implementing the unethical option in AsymmInfo have a higher SVO score, i.e., are more prosocial, than those doing the same in HighInfo ($p = 0.051$). This suggests that asymmetric information provides excuses enabling otherwise more prosocial decision-makers to act unethically.

Appendices

Appendix A

A.1 Additional material

A.1.1 Instructions

Welcome to the experiment

Thank you for your participation in this experiment. Please read the instructions carefully. For your participation today you will receive 5 €. During the experiment you will have the possibility to earn further money. Your additional payment will depend on your choices, the choices of other participants, as well as random events. Additionally, you will receive the earnings from the online part of the experiment at the end of today's experiment. After the experiment there will be a short questionnaire.

Please avoid any communication with your neighbors during the experiment. Switch off your mobile phone and remove everything you do not need for the experiment from the table. If you have any questions, please raise your hand and we will come to answer your questions at your seat.

Instructions

In this experiment, a participant decides in the role of **Participant A** how to distribute 10 € between himself and another randomly determined **Participant B**.

First, all participants decide **in the role** of **Participant A**. This means that you will decide how to distribute **10 €** between yourself and **Participant B**. You can allocate any amount between 0 € and 10 € in discrete intervals to Participant B. Participant B will receive this amount and you will receive the remaining amount. Your decisions will be kept anonymous and you will not know, neither during nor after the experiment, with which participant you interacted.

You will learn which role you have been assigned to only at the end of the experiment and after you have taken your decision. Half of the participants will be assigned the role of Participant A, while the other half of the participants will be assigned that of Participant B. That is, there are two possibilities:

1. You are selected as Participant A. This means: Your decision will be implemented. You will be randomly assigned to someone in the role of Participant B. You will receive 10 €, minus the amount you have allocated to Participant B. Accordingly, Participant B will receive the amount you allocated him.
2. You are selected as Participant B. This means: Your decision will not be implemented. You will be randomly assigned to someone in the role of Participant A. You will receive an amount of money according to the decision of Participant A.

Since, at the time of making your decision, you do not know whether you will be selected as Participant A or Participant B, please take your decision carefully.

After the experiment, a short questionnaire will follow. Then, the experiment will be concluded. We kindly ask you to stay seated. We will call participants individually and pay them in private. Do you have

further questions? Then, please raise your hand and we will come to answer your questions at your seat. Before the actual experiment starts, you will have to answer some questions of understanding.

A.1.2 Decision Screen

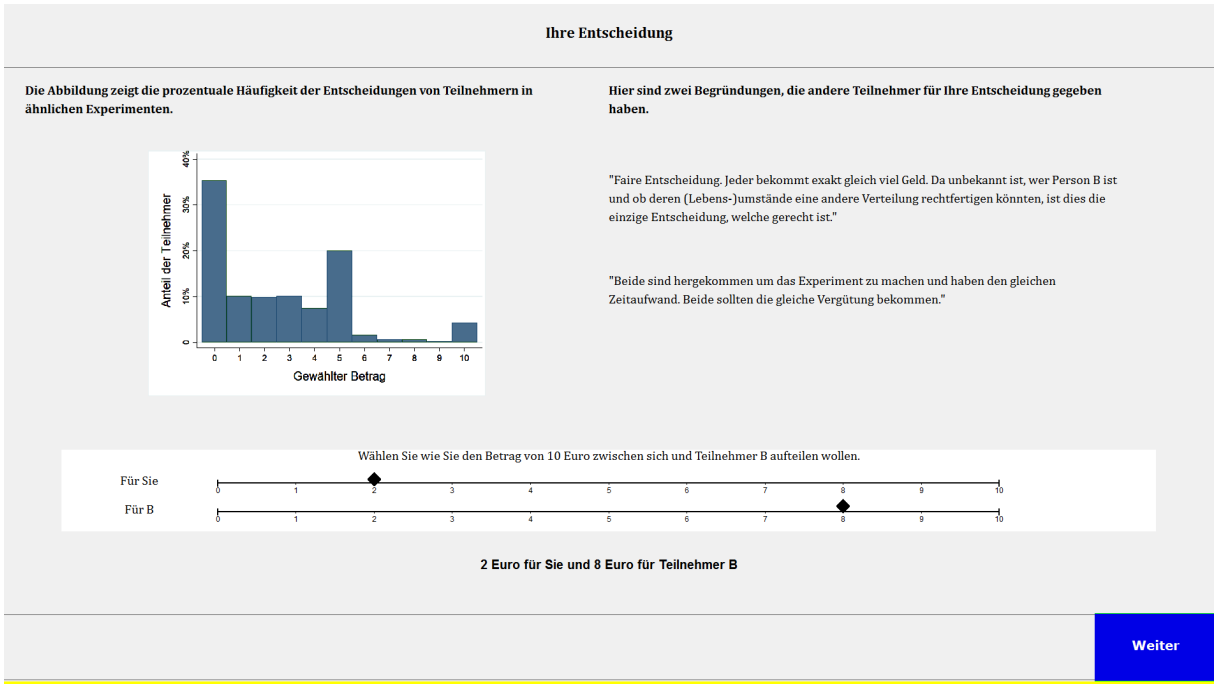


Figure A1: Dictator game decision screen

Note: The decision screen shows the empirical distribution of choices on the left. On the right side the two (positive or negative) narratives are listed. Below subjects take the dictator game decision.

A.1.3 Narrative Selection

The following table shows positive and negative narratives (translated from German) along with their average convincingness rating. Numbers 1-4 were selected for the POSITIVE condition and 5-8 for the NEGATIVE condition. Narratives were selected from all narratives of the first 3 sessions of the BASELINE condition, since the 4th session was run later to balance the number of participants in all conditions. More detailed information as well as the complete list of comments is available from the authors upon request.

Number	Positive Narratives	Convincingness
1	Both came here to participate in the experiment and spent the same amount of time here. Both should get the same payment.	6
2	An equal distribution of the money is only logical: Assuming everyone agrees on that, everyone will go home with 10 €. Everything else would be a mixture of greed and speculation.	6
3	Fair choice. Everyone gets exactly the same amount of money. Since it is unknown who Person B is and whether her life circumstances would justify another distribution, this is the only just decision.	6
4	I think that both participants should get the same amount of money. If it is unknown in advance whether you are A or B it is just smart to give 5 € to both.	6.3
	Negative Narratives	
5	Since the experiment is anonymous, I expect that everyone is looking for her own advantage. I don't know any of the other players and since the decision happens randomly anyway, I do not care about giving someone else money.	6
6	This way I get the highest payoff in case I am participant A. In case I am participant B, I have no influence on my payoff because of the assignment to role B.	5.6
7	Because I would like to have the money and saw in the statistic that others also decided this way. This made me have less scruples for allocating all the money to myself.	5.3
8	I allocated 10 € to myself, since this way I get the most money on average. As it is unclear how much I would get as participant B, I wanted to achieve the maximum profit in case I am participant A.	5.3

A.1.4 Additional psychological measures

Big 5 Questionnaire

This questionnaire is taken from Rammstedt and John (2007).

Instruction: How well do the following statements describe your personality?

I see myself as someone who …	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
… is reserved	(1)	(2)	(3)	(4)	(5)
… is generally trusting	(1)	(2)	(3)	(4)	(5)
… tends to be lazy	(1)	(2)	(3)	(4)	(5)
… is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)
… has few artistic interests	(1)	(2)	(3)	(4)	(5)
… is outgoing, sociable	(1)	(2)	(3)	(4)	(5)
… tends to find fault with others	(1)	(2)	(3)	(4)	(5)
… does a thorough job	(1)	(2)	(3)	(4)	(5)
… gets nervous easily	(1)	(2)	(3)	(4)	(5)
… has an active imagination	(1)	(2)	(3)	(4)	(5)

Context (In)dependence

This questionnaire is taken from Gollwitzer et al. (2006). The following is an English translation of the original questionnaire in German. Agreement to an item is measured on a 6 point Likert scale from "does not apply at all" to "fully applies".

Context dependence

1. My attitudes and opinions are often determined by the circumstances.
2. My behavior often depends on the people I am spending time with at that moment.
3. My decisions often depend on the temporary circumstances.
4. I behave very differently with different people.
5. My self-image depends overall on how other people perceive me.

Context independence

1. Once I have made a choice, I do not like to change it afterwards.
2. My self-image stays the same regardless of what others say about me.
3. I advocate for my own opinion regardless of the person with whom I am interacting.
4. I am the same person in different situations.
5. My attitudes and opinions hardly change, regardless of what happens in my life.

Moral disengagement

This questionnaire is taken from Bandura et al. (1996). We excluded the following categories: euphemistic language, attribution of blame and dehumanization, as they did not apply to our experimental framework. The following is an English translation of the version by Rothmund (unpublished), who validated the questionnaire in German. Agreement to an item was measured on a 6-point Likert scale from "do not agree at all" to "fully agree".

1. It is alright to beat someone who badmouths your family.
2. Arriving late is better than not coming at all.
3. It does not make sense to avoid flying to go on vacation for the sake of the environment, since everybody else does it as well.
4. It is okay to tell small lies because they don't really do any harm.
5. It is alright to lie to keep your friends out of trouble.
6. Given the million-dollar frauds of some mangers, one cannot be blamed for scrounging some office supplies.
7. It is not so bad to cheat on taxes, since everybody does it anyway.
8. One cannot be blamed for an offence, if he or she has been put under pressure by his or her friends.
9. Teasing someone does not really hurt them.
10. It is less bad to steal from the rich than from the poor.
11. A single person cannot be blamed for misbehaving, if everyone else does the same.
12. Managers cannot be blamed for layoffs, that is simply how business life works.
13. It is alright to leave some trash in the cinema hall, since it will be cleaned after the screenplay anyway.
14. The reason why poor people do not have money is that they are too lazy to work.

Moral identity

This questionnaire was originally developed by Aquino and Reed (2002). We use the German version validated by Rothmund and Gollwitzer (unpublished) and modified the list of attributes in the instructions. The following is an English translation of the material we used. Agreement to an item is measured on a 6-point Likert scale from "do not agree at all" to "fully agree".

Instructions: Below is a list of character attributes that might describe a person. The person with these attributes could be you, but also someone else.

Fair, generous, sympathetic, nice, and benign.

Imagine a person displaying exactly these character attributes. Imagine how this person would think, feel, and act. Once you have a precise image of this person, try to answer following questions.

1. It would make me feel good to be a person who has these characteristics.
2. Being someone who has these characteristics is an important part of who I am.
3. I would be ashamed to be a person who has these characteristics.
4. Having these characteristics is not really important to me.
5. I strongly desire to have these characteristics.
6. I often wear clothes that identify me as having these characteristics.
7. The types of things I do in my spare time (e.g., hobbies) clearly identify me as having these characteristics.
8. The kinds of books and magazines that I read identify me as having these characteristics.
9. The fact that I have these characteristics is conveyed to others by my membership in certain organizations.
10. I am actively involved in activities that convey to others that I have these characteristics.

A.1.5 Sessions

Session	Date (2018)	Treatment	Participants
1	May, 7	Baseline	22
2	May, 16	Baseline	24
3	May, 16	Baseline	28
4	May, 30	Positive	25
5	May, 30	Negative	22
6	May, 30	Positive	24
7	May, 30	Negative	26
8	June, 26	Positive	24
9	June, 26	Baseline	22
10	June, 26	Negative	25
11	June, 26	Negative	20
12	June, 26	Positive	18

Table A1: Session overview

Table A1 displays the date of the sessions and information about the treatment and the number of participant.

A.2 Theoretical framework

This section complements the "Behavioral Predictions" in the main text (Section 1.2.2) by providing formal definitions and derivations of the hypotheses. A decision maker chooses how much money to give to a recipient. A key component of this model is the belief about the externality of giving (Bénabou et al., 2018). We, first, describe the basic utility function of a decision maker; we, then, explain which role the externality plays; and, finally, discuss how narratives enter the model.

The utility function of a decision maker (DM) takes the following form:

$$U_i(g, e) = v(g, e) - c(g), \tag{A.1}$$

where g is the amount she decides to give, and e is the expected externality of giving, which we define below; $v(g, e)$ captures the overall valuation of giving, and $c(g)$ the costs of giving.[1] We set $e \in (0, 1)$ and assume $c(g)$ to be linear increasing in g. While $v(g, e)$ can take many functional forms, we assume concavity in g ($\frac{\partial v(g,e)}{\partial g} > 0$, $\frac{\partial^2 v(g,e)}{\partial g^2} < 0$). This assumption ensures an internal solution with an optimal amount of giving $g^*(e)$.

The externality. E is a binary measure of the presence of a positive externality, i.e., whether the recipient is deserving or it is appropriate to give in the situation at hand (see discussion in Section 1.2.2). If $E = 1$, there is a positive externality, while if $E = 0$, there is no such externality. A DM in our model does not know the value of E with certainty. Rather, she holds a prior belief (what we call perception above) about E with $e = P(E = 1)$. We assume that the marginal utility of giving is increasing in the expected externality e ($\frac{\partial v(g,e)/\partial g}{\partial e} > 0$). Following this assumption, a higher e leads to higher amounts of giving. Note that $v(g, e)$ can take on many different forms. In a setting like the standard dictator game the strong focal point at the equal split could be understood as a norm. Correspondingly, setting $v(g, e) = -\gamma(e)(\frac{1}{2} - g)^2$ in a dictator game with a pie size of 1, $\gamma(e)$ would capture the appropriateness to follow the norm, i.e., to split the pie equally (assuming $\frac{\partial \gamma}{\partial e} > 0$). Independently of the specific choice of v, our predictions hold.

Narratives. We model narratives as signals about E updating the prior belief of a DM, as in Bénabou et al. (2018). A positive narrative signals that $E = 1$, i.e., it is an argument or justification for there being a positive externality. A negative narrative, conversely, signals that $E = 0$. For simplicity, we take DMs to be standard Bayesian updaters. Other forms of updating are of course conceivable, but would introduce further degrees of freedom in the model. Moreover, as long as an alternative updating model leads to updating in the same direction for all priors and leads to different posteriors for different priors, the main intuitions of the model will hold. We assume narratives to be at least somewhat believable or convincing, which here means that the signal is correct more often than not. Hence, a DM will update in the direction of the signal.[2]

As an example, let us assume a signal structure as in Figure A2. If there is no externality $E = 0$, with probability $1 \geq c > \frac{1}{2}$ the correct signal, i.e. the negative narrative, is sent, and with $1 - c$

[1] Note that all factors influencing the utility of giving are captured by the first term. For the sake of simplicity, we do not consider how image concerns would alter the resulting trade-off.

[2] Note that Bénabou et al. (2018) formally define positive and negative narratives directly by their influence on beliefs. The signalling structure we use is based on an older version of their paper and leads to the same directional effect of narratives on actions.

the signal is wrong, i.e. the narrative is positive. The situation is reversed with a high externality ($E = 1$).

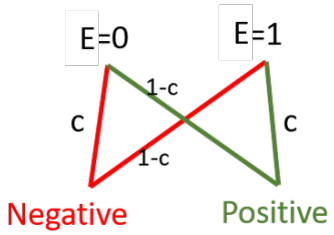


Figure A2: Exemplary signal structure

The posterior given a positive or negative signal is calculated as follows (with e being the prior probability of $E = 1$). Figure A3 provides a graphical representation.

$$P_{post}(E = 1|Positive) = \frac{P(Positive|E = 1)P_{prior}(E = 1)}{P(Positive)} = \frac{ce}{ce + (1 - c)(1 - e)}$$
$$P_{post}(E = 1|Negative) = \frac{P(Negative|E = 1)P_{prior}(E = 1)}{P(Negative)} = \frac{(1 - c)e}{(1 - c)e + c(1 - e)}$$

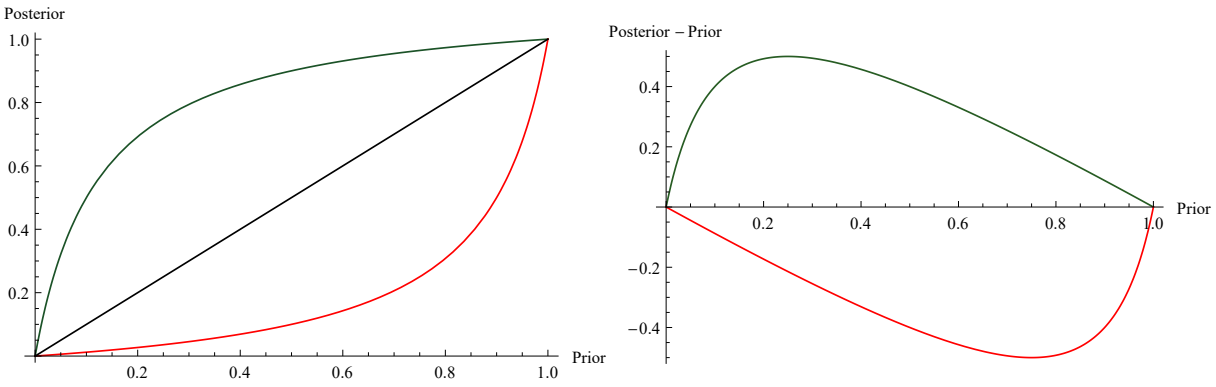


Figure A3: Posterior for given signal

Note: The left figure shows posterior beliefs as a function of prior beliefs and the right figure shows the corresponding difference between posterior and prior beliefs, both after receiving a positive (green, upper line) or negative signal (red, lower line), dependent on the prior belief. For these examples, we set $c = 0.9$. The black line on the left is the 45-degree line representing the case with no signal or no updating.

Given this signal structure, negative narratives lead to a downward shift in beliefs and positive narratives to an upward shift. That is, independent of the prior belief, the posterior belief is decreasing when receiving a negative narrative and increasing when receiving a positive narrative for the full range of beliefs. Since, as stated above, higher beliefs about e translate into higher amounts of giving, our first hypothesis follows directly.

Hypothesis A.1 *Positive narratives increase giving, while negative narratives decrease giving.*

Heterogeneity. We introduce heterogeneity by allowing diverging beliefs about E.[3] In fact, DMs in our model differ solely in their beliefs, which we bound to $e \in (0, 1)$. That is, all DMs in our model would act in the same way, i.e., choose to give the same amount, if they held the same belief. Modelling heterogeneity solely through beliefs offers us a concise way to introduce narratives as signals. We call DMs with low beliefs "selfish" types and those with high beliefs "prosocial" types.

While in our framework the direction of the effect of narratives is independent from prior beliefs, our setup predicts a different strength of the effect for different priors. In particular, extreme types (those with priors $\hat{e}$ close to 0 or close to 1) will not update strongly when receiving a signal close to their prior belief, whereas they will update strongly when receiving a contradicting signal (Figure A3).

Hypothesis A.2 *Positive narratives should have a stronger positive effect on more selfish types, while negative narratives should have a stronger negative effect on more prosocial types.*

A.2.1 Extension: Social Comparison

In this section, we provide an extension of our model and analyse the optimal giving behavior for a specification which captures our main results as well as our additional ones. Note that the goal is not to offer a general solution to the analytical problem here, but rather to show that the addition of a social comparison component can explain our findings.

The main idea is that narratives, on top of acting as a signal, provide a benchmark for social comparison. We introduce a social comparison component to the utility function which captures this idea. DMs gain from giving more than the narrator but this gain is decreasing for larger amounts, i.e., gains are concave in the positive difference between giving and the amount advocated for by the narrator. Conversely, giving a little less than the narrator leads to a large loss which marginally decreases for lower giving, i.e., it is convex in the difference of giving and the narrator's giving. The following specification reflects this and Figure A4 shows a potential social comparison function:

$$S(g, n) = \begin{cases} \mu(g - n)^{\alpha} & \text{if } g \geq n \\ -\mu(n - g)^{\alpha} & \text{if } g < n, \end{cases} \tag{A.2}$$

with $\alpha < 1$, where n determines the narrator's amount of giving and μ the weighting of the social comparison.

[3] Bénabou et al. (2018) hint at heterogeneity in priors, but consider common priors throughout the paper with heterogeneity between subjects stemming solely from different valuations of the externality.

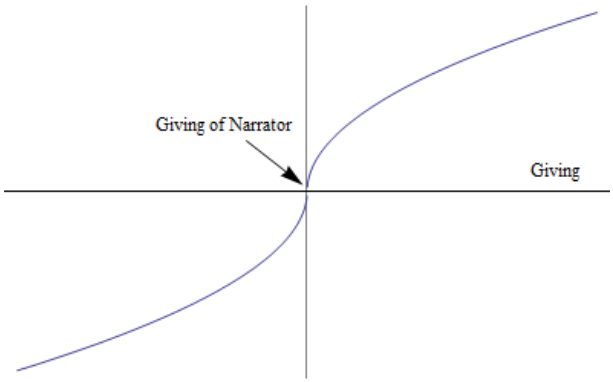


Figure A4: Social comparison function

Note: An example for a social comparison function with $\alpha = 0.5$.

We define the utility function of DMs as

$$U(g, e) = v(g, e) + S(g, n) - c(g), \tag{A.3}$$

where $S(g, n)$ describes a social comparison function as above. Importantly, the social comparison part is evoked by a narrative and disappears if there is no narrative (as in our BASELINE treatment).

The general solution of the above problem is not straightforward and might depend on the specifications of the value and social comparison functions. For tractability, we use a specification with a linear increasing giving function. This will lead to a step-function of giving for the POSITIVE as well as for the BASELINE treatment. We take the following specification:

$$U(g, e) = \begin{cases} 2eg + \mu(g - n)^{\frac{1}{2}} - g & \text{if } g \geq n \\ 2eg - \mu(n - g)^{\frac{1}{2}} - g & \text{if } g < n \end{cases} \tag{A.4}$$

with $g \in [0, 1]$ and e being the belief about the presence of the externality which can be influenced by a narrative as above.

In the example, we define the amount given by the narrator of a positive narrative as $n = 1$ (this reflects the natural, fair upper bound of giving 5 € in our experiment), and that of a narrator of a negative narrative as $n = 0$.

The resulting optimal giving functions in the three treatments are displayed in Figure A5 and formally presented below (e_{pos} and e_{neg} reflect the posterior beliefs about the presence of an externality in the positive and negative narrative treatment, respectively).

$$g^*(e, \mu)_{Baseline} = \begin{cases} 0 & \text{if } e < \frac{1}{2} \\ 1 & \text{if } e \geq \frac{1}{2} \end{cases} \tag{A.5}$$

$$g^*(e, \mu)_{Positive} = \begin{cases} 0 & \text{if } e_{pos}(c, e) < \frac{1-\mu}{2} \\ 1 & \text{if } e_{pos}(c, e) \geq \frac{1-\mu}{2} \end{cases} \tag{A.6}$$

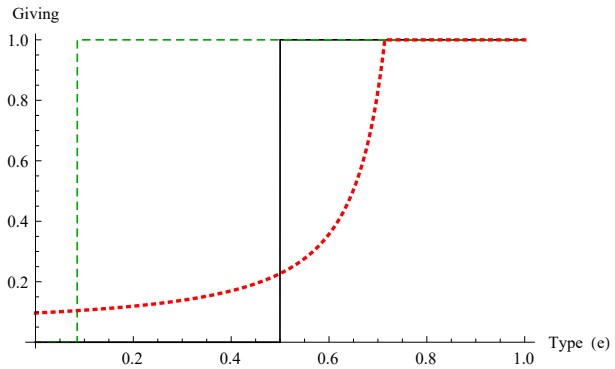


Figure A5: Predicted giving behavior

Note: The figure shows the predicted giving functions for the above specification in the BASELINE (black), NEGATIVE (red, dotted), and POSITIVE condition (green, dashed). Example parameters $\mu = 0.39$, $c = 0.83$.

$$g^*(e,\mu)_{Negative} = \min((\frac{\mu}{1-2e_{neg}})^2, 1) \tag{A.7}$$

The resulting predicted behavior according to the model shares key characteristics with our experimental results (see Figure 1.3 in Section 1.3). First, giving is higher in POSITIVE compared to BASELINE. Second, there is a differential effect for NEGATIVE narratives with prosocial types decreasing their giving and selfish types increasing their giving. Importantly, this example also captures the increase in equal splits, which is the action the narrator advocates for in POSITIVE, and the decrease of subjects giving nothing in NEGATIVE.

A.3 Additional analyses

In Table A2 we conduct multiple robustness checks. In the first column we control for the additional psychological measures. In column 2, we impose both lower and upper censoring. For interpretability of the interactions, we plot marginal effects as in the main text (see Figure A6). Column 3 introduces a quadratic term for types and interactions with the treatment conditions (see Figure A7 for the marginal effects). We normalize our type measure for this specification (in the graph, we show the most frequent non-normalized types as references). The pattern described in Section 1.3 remains qualitatively the same for all these alternative specifications. In column 4, we run a standard OLS regression. Coefficients have the same sign and significance level as in the Tobit regressions.

	Tobit controls	Tobit sessions	Tobit, upper and lower censoring	Tobit quadratic	OLS
POSITIVE	2.419***	1.884*	5.799***	6.856	1.468***
	(0.855)	(1,062)	(2.166)	(4.548)	(0.555)
NEGATIVE	2.635***	2.709**	5.494**	11.96**	1.313***
	(0.868)	(1.086)	(2.162)	(3.907)	(0.560)
Type	0.165***	0.163***	0.405***	38.78***	0.123***
	(0.0211)	(0.0211)	(0.0613)	(12.77)	(0.0133)
POSITIVE x type	-0.0580**	-0.0532**	-0.142**	-15.32	-0.0365**
	(0.0270)	(0.0269)	(0.0717)	(16.54)	(0.0187)
NEGATIVE x type	-0.0905***	-0.0918***	-0.194***	-36.41**	-0.0487***
	(0.0275)	(0.0275)	(0.0717)	(14.22)	(0.0188)
Type^2				-21.78**	
				(10.74)	
POSITIVE x type^2				8.518	
				(13.99)	
NEGATIVE x type^2				26.02**	
				(12.16)	
Constant	-3.685**	-3.5625*	-7.972***	-12.83***	-0.509
	(1.867)	(1.9015)	(1.815)	(3.558)	(0.397)
Controls	yes	yes	no	no	no
Session	no	yes	no	no	no
Observations	280	280	280	280	280
Pseudo R^2	0.144	0.1512	0.140	0.124	0.3647

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table A2: Robustness checks

Note: OLS and Tobit. The type measure corresponds to the SVO angle, POSITIVE and NEGATIVE conditions are introduced as dummies. We also include interaction terms between conditions and types. Controls include Context Dependence, Context Independence, Moral Identity Scale, Moral Disengagement, and the 11-item, Big-5 questionnaire. Session includes session dummies.

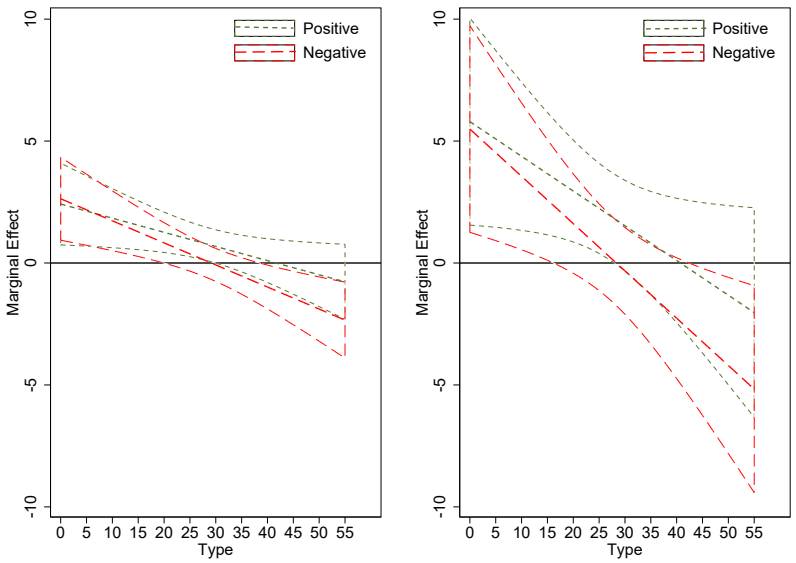


Figure A6: Marginal effects, Tobit.

Note: Tobit with controls left, Tobit with upper and lower censoring right. 95 %-confidence intervals

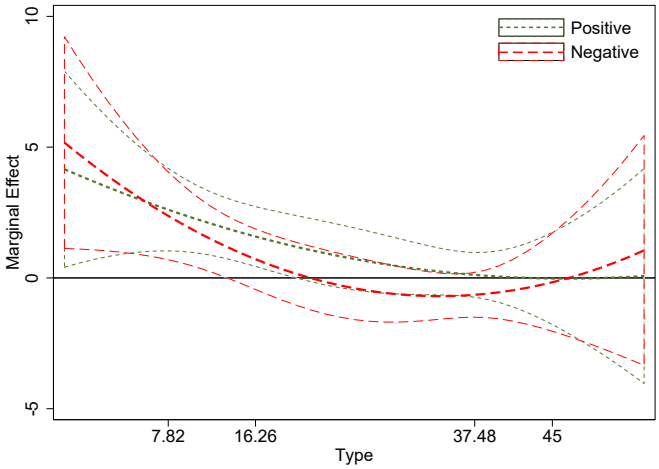


Figure A7: Marginal effects. Tobit with quadratic interaction term. 95 % confidence intervals

A.3.1 Analysis of additional psychological measures

In Table A3, we run the same analysis as in Section 1.3 using the additional psychological measures collected in the online pre-study. Both Moral Identity and Moral Disengagement have a strong and highly significant relationship with giving in the expected direction, i.e., positive and negative, respectively. However, they do not contribute significantly to the explanation of our treatment effects. Meaning that the NEGATIVE and POSITIVE condition do not affect subjects scoring differently on these scale in a different way. As to the complementary measures of Context Dependence and Independence, they do not significantly mediate our treatment effects. Meaning that the treatment conditions do not affect subjects who are more or less dependent from the context in making their decisions, as measured by these scales, differently.

	Moral identity	Moral disengagement	Context dependence	Context independence
POSITIVE	1.500	1.705	1.485	1.391
	(2.340)	(1.933)	(1.352)	(2.185)
measure	1.303***	-1.222**	-0.0443	0.116
	(0.401)	(0.489)	(0.243)	(0.412)
POSITIVE × measure	-0.270	-0.274	-0.211	-0.188
	(0.567)	(0.676)	(0.344)	(0.583)
NEGATIVE	0.308	0.823	-0.0235	0.495
	(2.399)	(2.053)	(1.402)	(2.171)
NEGATIVE × measure	-0.133	-0.248	0.0251	-0.117
	(0.581)	(0.735)	(0.352)	(0.587)
Constant	-2.913*	5.506***	2.329**	1.738
	(1.613)	(1.349)	(0.952)	(1.538)
Observations	280	280	280	280
Pseudo R^2	0.024	0.023	0.004	0.003

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table A3: Alternative measures

Note: Tobit regression with censoring at 0. Giving on treatment and stated measures as well as the interaction term.

A.3.2 Probit regressions

	give 5	give 0
POSITIVE	1.559**	-0.975**
	(0.653)	(0.471)
Type	0.0820***	-0.0825***
	(0.0167)	(0.0131)
POSITIVE x type	-0.0386*	0.0204
	(0.0198)	(0.0186)
NEGATIVE	1.020	-1.230***
	(0.694)	(0.457)
NEGATIVE x type	-0.0328	0.0491***
	(0.0209)	(0.0166)
Constant	-2.705***	1.617***
	(0.568)	(0.352)
Observations	280	280
Pseudo R^2	0.213	0.275

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table A4: Probit regressions

Note: Probit regression. Dependent variable is a dummy of giving 5 in the first column and a dummy of giving 0 in the second column. Independent variables are treatment conditions, type, and interaction terms.

A.3.3 Feelings

In Table A5, we regress the measures of feelings we collected after subjects' choice in the dictator game. In all columns, we regress a specific measure on dummies for treatment conditions, the amount a subject gave, her SVO angle and an interaction term between the latter and the treatment conditions. The first two columns refer to general feelings of happiness and contentment (how happy/contented do you feel at the moment?), which are rather stable. The last four columns refer to feelings regarding a subject's choice in the dictator game. Guilt and shame decrease in the amount a subject gives. However, the presence of narratives in our treatment conditions does not substantially alter this relationship.

	Happiness	Content	Guilt	Contentment	Shame	Excited
Constant	4.137***	3.854***	2.440***	4.169***	2.089***	2.598***
	(0.319)	(0.331)	(0.264)	(0.261)	(0.229)	(0.326)
Positive	0.694	0.756	0.455	0.318	0.240	0.553
	(0.451)	(0.468)	(0.373)	(0.369)	(0.323)	(0.461)
Negative	0.651	1.034*	−0.127	0.454	0.246	−0.027
	(0.454)	(0.470)	(0.376)	(0.371)	(0.325)	(0.464)
Type	0.013	0.017	0.012	0.018	0.005	0.001
	(0.012)	(0.013)	(0.010)	(0.010)	(0.009)	(0.013)
Give	−0.003	0.040	−0.309***	0.001	−0.213***	0.032
	(0.048)	(0.050)	(0.040)	(0.040)	(0.035)	(0.050)
Positive × Type	−0.012	−0.019	−0.014	−0.012	−0.008	−0.018
	(0.015)	(0.016)	(0.012)	(0.012)	(0.011)	(0.015)
Negative × Type	−0.017	−0.023	0.008	−0.017	−0.002	0.000
	(0.015)	(0.016)	(0.013)	(0.012)	(0.011)	(0.016)
Adj. R^2	−0.004	0.009	0.210	−0.005	0.162	−0.012
Num. obs.	280	280	280	280	280	280

Standard errors in parentheses
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A5: Regression analysis for measures of feelings

Note OLS of stated feeling on treatment, type, and the interaction term. The first two columns refer to general feelings, the last 4 columns refer to feelings specific to the choice.

Appendix B

B.1 Additional material

Instructions translated from German.

Welcome to this experiment!

Thank you for your participation in this experiment. Please read the instructions carefully. For your participation today you will receive **4 Euro**. During the experiment you will have the possibility to earn further money. Your additional payment will depend on your choices, the choices of other participants, as well as random events.

Please switch off your mobile phone and avoid any communication with your neighbors. We kindly ask you to read the following instruction carefully. If you have any questions, please raise your hand and we will come to answer your questions at your seat.

This experiment has two parts. You will first work on Part 1. When you will be finished with Part 1, you will receive the instruction of Part 2.

Part 1 In this part you have to work on a given number of tasks just as all other participants to today's experiment.

These tasks consist of multiple sliders that will appear on your screen. Each slider has to be moved from its initial location to the "target" location. When the target location is reached, the value will be displayed in green.

All participants have to complete 12 such sliders. Each participant will get **10 tokens** for the completion of these tasks.

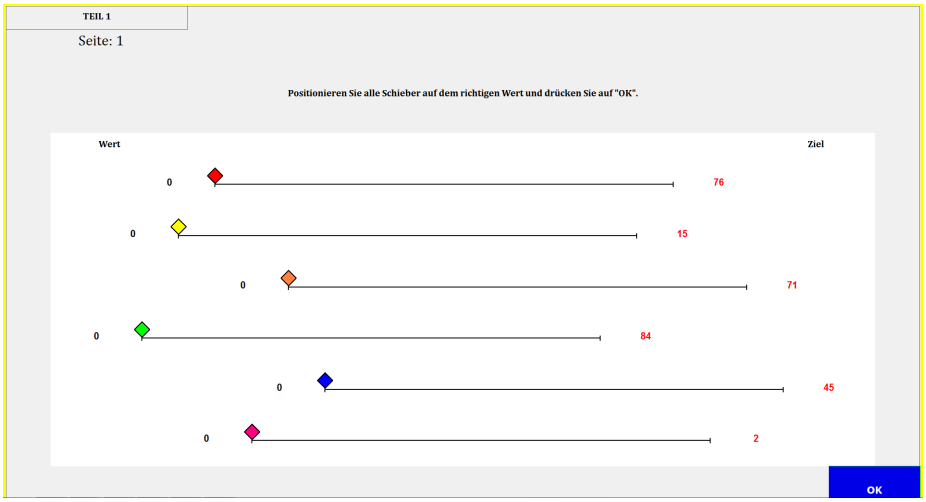


Figure B1: Slider Task

Instructions of Allocators:

Part 2 In this part of the experiment, you have been paired with another participant. You have been randomly assigned to the role of **Participant A**. The other participant has been randomly assigned to the role of **Participant B**.

Your task: Both participants have earned 10 tokens in Part 1. Hence, together you obtained 20 tokens. As Participant A you have to distribute these **20 tokens** between yourself and Participants B. You can take up to 8(2) tokens from Participant B and you can give up to 8(2) of your own tokens to Participant B.

Task of Participant B: Participant B decides at which price the 20 tokens are sold. He can set a price of 0.5 € per token (20 tokens = 10€) or a price of 0.25 € per token (20 tokens = 5 €), whereby he then will get an additional compensation of 5 € just for himself.

On the next screen you will see the options of Participant B in detail.

These are not your instructions! These are the complete instructions of Participant B.

Participant A is going to distribute the 20 tokens between himself and you. He will, hence, decide how many tokens he wants to keep and how many he wants to give to you. As Participant B you have to need to decide what is the value of the tokens:

Option A: Each token will be worth 0,50 € (20 tokens = 10 €).

Option B: Each token will be worth 0,25€ (20 Taler = 5 Euro)., and as compensation you will receive 5 € just for yourself.

Participant A will know only after his decision which of the two options you have chosen. Likewise, you will get to know how Participant A distributed the tokens only after your decision.

Please note that the instructions to the seller do not specify how many tokens you can distribute as Participant A.

You can now go back to your own instructions by clicking on the box on the right ("Part2"). Comprehension questions will now follow.

Instructions of Sellers:

Part 2 In this part of the experiment, you have been paired with another participant. You have been randomly assigned to the role of Participant B. The other participant has been randomly assigned to the role of Participant A.

Both participants have earned 10 tokens in Part 1. Hence, together you obtained 20 tokens.

Participant A is going to distribute the 20 tokens between himself and you. He will, hence, decide how many tokens he wants to keep and how many he wants to give to you. As Participant B you have to need to decide what is the value of the tokens:

Option A: Each token will be worth 0,50 € (20 tokens = 10 €).

Option B: Each token will be worth 0,25€ (20 Taler = 5 Euro)., and as compensation you will receive 5 € just for yourself.

Participant A will know only after his decision which of the two options you have chosen. Likewise, you will get to know how Participant A distributed the tokens only after your decision.

Appendix C

C.1 Additional material

C.1.1 Instructions online experiment

These are the instructions used in the online experiment (first part). The original text was in German and is available upon request.

Welcome Welcome and thank you for your participation in this study.

This study is composed by two parts, today's online part (first part) and a part in the premises of the Cologne Laboratory for Economic Research (second part). This online part will take ca. 30-45 minutes. Please be aware that you need to complete this online part to take part in the second part. You will receive an email to remind you of this.

Please complete this part all together, undisturbed and concentrated. If possible, please use a computer or a tablet. Please avoid other disturbances and complete this study alone. We reserve ourselves the right to exclude participants who do not complete the study carefully from the experiment.

All your decision will be used only for scientific purposes and for determining your payment.

You will get a fix payment of 8 € for participation after you have completed both parts of the study. You have the opportunity to earn further amounts of money during this first part as well as during the second part. For this reason, please read the following instruction carefully.

The further earnings from the online part as well as the earnings from the part in the Cologne Laboratory for Economic Research and the fixed payment of 8€ will be paid out after the part in the Cologne Laboratory.

Please click on the arrow below to start with the study.

Code

In order to guarantee your payment, you have to generate a code below. You will generate the exact same code in the second part of the experiment. We will use your code to execute your payment anonymously.

Please insert your personal code in lower-case letters and without accents or other special symbols.

The code is composed by the following components:

SECOND letter of your own name

FIRST letter of your mother's name (if unknown insert "*")

FIRST letter of your father's name (if unknown insert "*")

SECOND letter of the name of your birthplace (if unknown insert "*")

Day of your birthday (e.g., 15 for 15/07 or 08 for 08/03)

Please type in the code in small letters and without accents or other special symbols.

Please do not use any umlaut. Write a instead of ä, o instead of ö and u instead of ü.

Example: Max Mustermann, son of Lisa and Paul, born in Bonn on the 27/04 the resulting code would be alpo27.

This online part is composed by three parts. You will obtain the corresponding instructions before each part and then complete that part.

Part 1 (personal norms)

In this part of the study, you will read the description of different situations. In each situation there is one person who has to make a choice between different actions.

After you read the description of each situation, you have to evaluate the different actions amongst which the person in that situation can choose from. For each action evaluate according to your own opinion and independently form the opinion of others, whether it is appropriate or not choose it. By "appropriate" behavior, it is meant behavior that you personally consider to be "correct" or "moral". The standard is, hence, your personal opinion, independently from the opinion of others.

We kindly ask you to answer as precisely as possible with your own honest opinion. There is no right or wrong answer; you will not get any additional payment for your answers in this part.

Overall there are four different situations for which you have to evaluate the possible actions. To show you how the different actions can be evaluated we now give you an example.

Example

Person A is sitting in a cafe near the university. Person A notices that another person has left his wallet on the table. Person A has to decide what to do. Person A has to choose from four possible actions:

- Take the wallet and keep it;
- Ask other guests, if the wallet belongs to one of them;
- Leave the wallet there;
- Give the wallet to the manager of the cafe.

For each action evaluate according to your own personal opinion and independently from the opinion of others, whether it is appropriate or not to choose it. By "appropriate" behavior, it is meant behavior that you personally consider to be "correct" or "moral".

You can choose from a scale with six points

- Very inappropriate

- Inappropriate
- Rather inappropriate
- Rather appropriate
- Appropriate
- Very appropriate

You will evaluate the actions using a table. To evaluate the behavior you have to mark the corresponding option. Please give an evaluation for each of the actions.

Assume, for example, that you evaluate

- Taking and keeping the wallet as very inappropriate,
- Asking other guests, if the wallet belongs to them as rather appropriate,
- Leaving the wallet there as rather inappropriate,
- Giving the wallet to the manager of the cafe as very appropriate

You would insert following evaluations.

After clicking on "next" the description of the actual situations that you have to evaluate will follow.

Description of the situations

1 Dictator game

In a study conducted at the economic laboratory, Person A is randomly matched with another participant, Person B. The matching is anonymous, hence no participant will ever learn about the identity of the other participants.

In this study, Person A takes a decision. Person B knows which decision Person A has to take. Person B also knows which consequences this decision has for the monetary payment and will know which decision Person A has taken.

Person A's decision

Person A gets 10 € at the beginning of the task. Person A can, then, give any amount of this 10 € to Person B.

Person A can, for example, give 0€ to Person B. Person A would get 10€ and Person B 0 €. Person A could also give 10 €. Person A would, then, get 0 € and Person B 10 €. Similarly, Person A could give 1€, 2€, 3€... or 9€.

At first, both participants will take a decision in the role of Person A. This means that both will indicate how many Euros they would give to Person B, in case you would be assigned to the role of Person A. After both participants have taken their decision, they will learn who was assigned to the role of Person A and Person B. Both participants will be paid according to the assignment of roles and the decisions taken.

Please evaluate the possible actions of Person A.

2 Dictator game with tax

In a study conducted at the economic laboratory, Person A is randomly matched to another participant, Person B. The matching is anonymous, hence no participant will ever learn about the identity of the other participants.

In this study, Person A takes a decision. Person B knows which decision Person A has to take. Person B also knows which consequences this decision has for the monetary payment and will know which decision Person A has taken.

Person A's decision

Person A gets 12 € at the beginning of the task. Person B gets 0 €. Person A can, then, send an amount of this 12 € to Person B. Person B gets 0.90 € for each 1.50 € Person A sends to him. Hence, 40% of the amount sent gets lost.

Person A can, for example, give 0€ to Person B. Person A would get 12 € and Person B 0 €. Person A could also give 12 €. Person A would, then, get 0 € and Person B 7.20 €. Similarly, Person A could give 1.50 €, 3 €, 4.50 €, ... or 10.50 €. You can find an overview of the possible actions and the corresponding earnings here:

A sends	0 €	1.50 €	3 €	4.50 €	6 €	7.50 €	9 €	10.50 €	12 €
hence Person A and Person B:									
A earns	12 €	10.50 €	9 €	7.50 €	6 €	4.50 €	3 €	1.50 €	0 €
B earns	0 €	0.90 €	1.80 €	2.70 €	3.60 €	4.50 €	5.40 €	6.30 €	7.20 €

At first, both participants take a decision in the role of Person A. This means that both have to indicate how many Euros they would send to Person B, in case they would be assigned to the role of Person A. After both participants have taken their decision, they will learn who was assigned to the role of Person A and Person B. Both participants are paid according to the assignment of roles and the decisions taken.

Please evaluate the possible actions of Person A.

3 Third-party punishment game

In a study conducted at the economic laboratory, Person C is randomly matched to two another participants, Person A and Person B. The matching is anonymous, hence no participant will ever learn about the identity of the other participants.

In this study, Person C and Person A take a decision. All three participants know which decision Person A and Person C have to take. They also know which consequences these decisions have for the monetary payment and will know which decision have been taken.

Person A gets 10 € at the beginning of the task. Person B gets 0 €. Person C gets 5 €.

Persons A's decision

Person A can give Person B 0€, 2€, or 5€ of his 10€. Person A could give Person B 0€. Then, Person A would get 10 € and Person B 0 €. Person A could also give Person B 5 €. If Person A would give 5 €, then he would earn 5 € and Person B would earn 5 € as well. If Person A would give 2 €, then he would earn 8 € and Person B 2 €.

Person C's decision

Person C can assign reduction points to Person A depending on his decision. Person C can assign 0, 1 or 2 reduction points to Person A. The earnings of Person C are reduced by 1 € for each reduction point and Person A's by 3 €. The earning of Person A cannot, however, go below 0 €, this means that his earnings can be reduced only until 0 €. The assignment of reduction points has no consequence for Person B.

If Person C would, for example, assign 0 reduction points, then neither the earnings of Person C nor those of Person A would be reduced. If Person C would assign 1 reduction points, then his earnings would be reduced by 1 € and those of Person A by 3 €. If Person C would assign 2 reduction points, then his earnings would be reduced by 2 € and those of Person A by 6 €.

Person C has to indicate how many reduction points he would assign Person A for each of his possible decisions (0 €, 2 €, or 5 €). Only the decision of Person C that corresponds to the actual decision of Person A is implemented.

Example: Person A gives 2 € to Person B. Person C indicated that in this case he would assign him 1 reduction point. Then, Person A would get a deduction of 3 € and Person C of 1 €. In this case Person A would hence get (8 € - 3 € =) 5 €, Person B 2 € and Person C (5 € - 1 € =) 4 €.

One of the participants is assigned to the role of Person A. The other two take, at first, both a decision in the role of Person C. Both indicate how many reduction points they would assign to Person A in case they were Person C. The other participants will learn only after their decision who was assigned to the role of Person B and who to that of Person C. Participants are paid according to the assignment of roles and the decisions taken.

(1) Assume Person A decides to give Person B 0 €. He, hence, keeps 10 € while Person B gets 0 €. Please evaluate the possible actions of Person C. (2) Assume Person A decides to give Person B 2€. He, hence, keeps 8€ while Person B gets 2 €. Please evaluate the possible actions of Person C. (3) Assume Person A decides to give Person B 5 €. He, hence, keeps 5 € while Person B gets 5 €. Please evaluate the possible actions of Person C.

4 Ultimatum game

In a study conducted at the economic laboratory, Person A is randomly matched to another participant, Person B. The assignment is anonymous, hence no participant will ever learn about the identity of the other participants.

In this study, Person A and Person B take decisions simultaneously. Both know which decision the other has to take. They also know which consequences this decision has for the monetary payment and will know in the end which decision the other has taken. Here is a description of the decision of Person A and Person B.

Person A gets 10 € at the beginning of the task. Person B gets 0 €. Person A and Person B then take a decision simultaneously.

Person A's decision

Person A can propose any amount of the 10 € to Person B. Person A, hence, decides how much of the 10 € he wants to propose to Person B.

Person B's decision

Person B decides which proposals he is ready to accept. The two participants get the stipulated amounts only if Person B accepts the offer. If he rejects the offer, both get 0 €.

To that, Person B chooses an amount between 0 € and 10 €. This amount is the lowest proposal that Person B is still ready to accept. All proposal that are equal to or higher than this amount are accepted by Person B. All proposals that are lower than this amount are rejected by Person B.

Person A does not know how much money Person B will at least accept at the point of his decision, since the decisions are taken simultaneously. Similarly, Person B does not know how much money Person A will actually propose at the point of his decision.

For example, Person B could only accept proposals above 2 €. Proposals of 0 € and 1 € would be rejected. All other proposals will be accepted. Person B could also only accept proposals above 8 €. Then, only proposals of 8 €, 9 € or 10 € would be accepted and all other offers would be rejected.

Please evaluate the possible actions of Person B.

Your task For each action evaluate according to your own personal opinion and independently from the opinion of others, whether it is appropriate or not to choose it. By "appropriate" behavior, it is meant behavior that you personally consider to be "correct" or "moral".

Part 2 (Social norms)

In this part of the study you will read the description of different situations. In each situation there is one person who has to make a choice between different actions.

After you read the description of each situation, you have to evaluate the different actions amongst which the person in the situation can choose from. For each action evaluate according to the opinion of the society and independently form your own opinion, whether it is appropriate or not choose it. By "appropriate" behavior, it is meant the behavior that you consider most people would agree upon as being "correct" or "moral". The standard is, hence, not your personal opinion, but your assessment of the opinion of the society. We kindly ask you to answer as precisely as possible.

In this part, you can earn up to 12 € on top of your participation fee of 8 € depending on your answers. The answers of the other participants will influence your payment in this part.

At the end of the study, we will determine for each action in each situation which answer most of the other participants gave. You will obtain 0.30 € for each action for which you give the same answer as most of the other participants.

Your payment is determined in the following way: you will evaluate the possible actions a person according to the opinion of the society in 4 different situations. For each action in each situation the following holds: if your evaluation is exactly the same as the answer of most of the other participants, you will earn money. For each correspondence you get 0.30 €. This means that you can earn up to 12 € in addition to the fixed participation fee of 8 €. If, on the contrary, you never give the same answer as most of the other participants, then you will earn no money in this task. If, for example, you give the most frequent answer for 10 actions, you get 3€ for this task.

Note: only the answers of other participants in this part count. All other participants have received the same instructions. Also, they get 0.30 € for each action for which they give the

same answer as most other participants.

Overall there are four different situations for which you have to evaluate the possible actions. To show you how the different actions can be evaluated, we now give you an example.

Example

Person A is sitting in a cafe near the university. Person A notices that another person has left his wallet on the table. Person A has to decide what to do. Person A has to choose from four possible actions:

- Take the wallet and keep it;
- Ask other guests, if the wallet belongs to one of them;
- Leave the wallet there;
- Give the wallet to the manager of the cafe.

For each action evaluate according to the opinion of the society and independently form your own opinion, whether it is appropriate or not choose it. By "appropriate" behavior, it is meant the behavior that you consider most people would agree upon as being "correct" or "moral". Note: you earn 0.30 € for each action for which your answer matches the most frequent answer of the other participants in this second part.

You can choose from a scale with six points

- Very inappropriate
- Inappropriate
- Rather inappropriate
- Rather appropriate
- Appropriate
- Very appropriate

You will evaluate the actions using a table. To evaluate the behavior you have to mark the corresponding option. Please give an evaluation for each of the actions.

Assume, for example, that you evaluate

- Taking and keeping the wallet as very inappropriate,
- Asking other guests, if the wallet belongs to them as rather inappropriate,
- Leaving the wallet there as rather inappropriate,
- Giving the wallet to the manager of the cafe as very appropriate.

You would insert following evaluations.

Assume the other participants give the following evaluations. The table below shows the percentage of participants that gave a given evaluation for each action. Obviously, you will not get this information in the actual situations. This example should help you to understand how you can earn additional money.

Aktion	Sehr unangemessen	Unangemessen	Eher unangemessen	Eher angemessen	Angemessen	Sehr angemessen
den Geldbeutel nehmen und behalten	50%	30%	15%	5%	0%	0%
andere Gäste fragen, ob ihnen der Geldbeutel gehört	0%	5%	10%	40%	25%	20%
den Geldbeutel liegen lassen	15%	20%	40%	20%	0%	5%
den Geldbeutel dem Cafébetreiber	0%	0%	0%	10%	30%	60%

How much additional money (in Cent) would you get for this situation? (If, for example, the correct answer is 1.5 €, then write "150".)

After you have answered this question, the description of the actual situation that you have to evaluate will follow.

Description of the situations

Repetition of the description of the situations (see above).

Your task For each action evaluate according to the opinion of the society and independently form your own opinion, whether it is appropriate or not choose it. By "appropriate" behavior, it is meant the behavior that you consider most people would agree upon as being "correct" or "moral". Note: you earn 0.30 € for each action for which your answer matches the most frequent answer of the other participants in this second part.

Part 3

This is the last part of this online part. Please fill in this questionnaire.

Which gender do you belong to?

Are you studying?

If yes, in which major?

How old are you?

Do you have any siblings?

What is your favorite movie?

What is your favorite food?

C.1.2 Instructions laboratory experiment

These are the instructions used in the laboratory experiment (second part). The original text was in German and is available upon request.

C.1.3 Instructions

Welcome Welcome to the second part of the study!

Today, you will take part to the second part of this study. You have already completed the first part online. You will be able to earn money in addition to the fixed amount of 8€and the amount you earned during the online study.

The level of this additional amount depends from your decision, the decision of other participants and chance. Thus, please read the instructions carefully.

Please avoid any conversation with your neighbor. Switch off your mobile phone and remove any item you do not need for the study from your table. In case you have questions, raise your hand and we will answer your question at your seat.

Code

Please insert your code from the online study below so that we can carry out your payment correctly at the end of the study.

Reminder: The code is composed by the following components:

SECOND letter of you own name

FIRST letter of your mother's name (if unknown insert "*")

FIRST letter of your father's name (if unknown insert "*")

SECOND name of your birthplace (if unknown insert "*")

Day of your birthday (e.g., 15 for 15/07 or 08 for 08/03)

Please type in the code in small letters and without accents or other special symbols.

Please do not any umlaut. Write a instead of ä, o instead of ö and u instead of ü.

Instructions before the games

Private treatment

Today's study is composed by four tasks. The tasks will be presented in a random order. You will receive instruction before each task and you will then work on the task.

In these tasks you will be matched with other participants. You and other participants will take decisions during these tasks. You can be matched with each participant only once – you will not be assigned to the same participant in two different tasks.

One of the tasks will be randomly selected for the payment of today's study. Since you will not know which task will be chosen until the end of the study, please go through the tasks carefully. At the end of the session we will pay you out the whole sum you earned during this study (8€ participation fee as well as the money from the online study and your payment from today's session) in cash.

Social treatment

Today's study is composed by four tasks. The tasks will be presented in a random order. You will receive instruction before each task and you will then work on the task.

In these tasks you will be matched with other participants. You and other participants will take decisions during these tasks. You can be matched with each participant only once – you will not be assigned to the same participant in two different tasks.

One of the tasks will be randomly selected for the payment of today's study. Since you will not know which task will be chosen until the end of the study, please go through the tasks carefully. At the end of the session we will pay you out the whole sum you earned during this study (8€ participation fee as well as the money from the online study and your payment from today's session) in cash.

When all participants in the session have completed the tasks everyone will have to stand up (so that all participants can hear and see each other). An assistant will call the participants one after the other. Each participant will have to say his name and tell the other participants which choices he made in the tasks. Thereto, a text will be displayed on your screen and you will have to read it verbatim. This means that all other participants will know your name and all the choices you have made in the tasks.

Games

Dictator game

In this task, you will be matched randomly to another participant. You will not know neither before nor after the study who the other participant has been.

You and the other participant will be assigned one of two roles: Person A or Person B.

Person A's decision

Person A gets 10 € at the beginning of the task. Person A can, then, give any amount of this 10 € to Person B. Person A can, for example, give 0€ to Person B. Person A would get 10 € and Person B 0 €. Person A could also give 10 €. Person A would, then, get 0 € and Person B 10 €. Similarly, Person A could give 1€, 2 €, 3 €, ... or 9 €.

At first, you and the other participant will both take a decision in the role of Person A. This means that you will indicate how many Euros you would give to Person B, in case you would be assigned to the role of Person A. Both of you will learn which role you have been assigned (Person A or Person B) only at the end of the study. The earnings of both participants result from the assignment of roles and the decision taken by Person A.

Before you take your decision on the next page, please answer the following two questions.

1. How much does Person A earn, if Person A gives 3 € to Person B?
2. How much does Person A earn, if Person A gives 1 € to Person B?

Ultimatum game

In this task, you will be matched randomly to another participant. You will not know neither before nor after the study who the other participant has been.

One participant is randomly assigned to the role of Person A and the other to the role of Person B.

Person A gets 10 € at the beginning of the task. Person B gets 0 €. Person A and Person B then take a decision simultaneously.

Person A's decision

Person A can propose any amount of the 10 € to Person B. Person A, hence, decides how much of the 10 € he wants to propose to Person B.

Person B's decision

Person B decides which proposals he is ready to accept. The two participants get the stipulated amounts only if Person B accepts the offer. If he rejects the offer, both get 0 €.

To that, Person B chooses an amount between 0 € and 10 €. This amount is the lowest proposal that Person B is just ready to accept. All proposal that are equal to or higher than this amount are accepted by Person B. All proposals that are lower than this amount are rejected by Person B.

Person A does not know how much money Person B will at least accept at the point of his decision. Similarly, Person B does not know how much money Person A will actually propose at the point of his decision.

For example, Person B could only accept proposals above 2 €. Proposals of 0 € and 1 € would be rejected. All other proposals will be accepted. Person B could also only accept proposals above 8 €. Then, only proposals of 8 €, 9 € or 10 € would be accepted and all other offers would be rejected.

Dictator

You were assigned to the role of Person A. The other participant was assigned to the role of Person B.

Before you take your decision on the next page, please answer the following two questions.

1. How much would Person A and Person B earn, if Person A offers Person B 4 € and Person B accepts the offer?
2. How much would Person A and Person B earn, if Person A offers Person B 2 € and Person B ...
    a ... accepts the offer?
    b ... rejects the offer?

Receiver

You were assigned to the role of Person A. The other participant was assigned to the role of Person B.

Before you take your decision on the next page, please answer the following two questions.

1. How much would Person A and Person B earn, if Person A offers Person B 4 € and Person B accepts the offer?
2. How much would Person A and Person B earn, if Person A offers Person B 2 € and Person B ...
   a ... accepts the offer?
   b ... rejects the offer?

Dictator game with tax

In this task, you will be matched randomly to another participant. You will not know neither before nor after the study who the other participant has been.

You and the other participant will be assigned one of two roles: Person A or Person B.

Person A's decision

Person A gets 1 € at the beginning of the task. Person B gets 0 €. Person A can, then, send an amount of this 12 € to Person B. Person B gets 0.90 € for each 1.50 € Person A sends to him. Hence, 40% of the amount sent gets lost.

Person A can, for example, give 0 € to Person B. Person A would get 12 € and Person B 0 €. Person A could also give 12 €. Person A would, then, get 0 € and Person B 7.20 €. Similarly, Person A could give 1.50 €, 3 €, 4.50 €, ... or 10.50 €. You can find an overview of the possible actions and the corresponding earnings here:

A sends	0 €	1.50 €	3 €	4.50 €	6 €	7.50 €	9 €	10.50 €	12 €
hence Person A and Person B:									
A earns	12 €	10.50 €	9 €	7.50 €	6 €	4.50 €	3 €	1.50 €	0 €
B earns	0 €	0.90 €	1.80 €	2.70 €	3.60 €	4.50 €	5.40 €	6.30 €	7.20 €

At first, you and the other participant will both take a decision in the role of Person A. This means that you will indicate how many Euros you would send to Person B, in case you would be assigned to the role of Person A. Both of you will learn which role you have been assigned (Person A or Person B) only at the end of the study. The earnings of both participants result from the assignment of roles and the decision taken by Person A.

Before you take your decision on the next page, please answer the following two questions.

1. How much does Person A earn, if Person A sends 1.50 € to Person B?
2. How much does Person A earn, if Person A sends 9 € to Person B?

Third-party punishment game

In this task, you will be matched randomly to another participant. You will not know neither before nor after the study who the other participant has been.

One participant will be assigned to the role of Person A, another one to the role of Person B and another one to the role of Person C. Person A gets 10 € at the beginning of the task. Person B gets 0 €. Person C gets 5 €.

Persons A's decision

Person A can give Person B 0 €, 2 €, or 5 €. Person A could give Person B 0 €. Then, Person A would get 10 € and Person B 0 €. Person A could also give Person B 5 €. If Person A would give 5 €, then he would earn 5 € and Person B would earn 5 € as well. If Person A would give 2 €, then he would earn 8 € and Person B 2 €.

Person C's decision

Person C can assign reduction points to Person A depending on his decision. Person C can assign 0,1 or 2 reduction points to Person A. The earnings of Person C is reduced by 1 € for each reduction point and Person A's is reduced by 3 €. The earning of Person A cannot, however, go below 0 €, this means that his earnings can be reduced only until 0 €. The assignment of reduction points has no consequence for Person B.

If Person C would, for example, assign 0 reduction points, then neither the earnings of Person C nor those of Person A would be reduced. If Person C would assign 1 reduction points, then his earnings would be reduced by 1€ and those of Person A by 3 €. If Person C would assign 2 reduction points, then his earnings would be reduced by 2€ and those of Person A by 6 €.

Example: Person A gives 2 € to Person B. Person C indicated that in this case he would assign him 1 reduction point. Then, Person A would be deducted 3 € and Person C 1 €. In this case Person A would hence get (8 € - 3 € =) 5 €, Person B 2 € and Person C (5 € - 1 € =) 4 €.

Person B does not make any decision in this task.

Dictator

You have been assigned to the role of Person A.

The other two participants were assigned to the role of Person A and Person C. At first, both of the other participants will take a decision in the role of Person C. Both indicate how many reduction points they would assign to you (Person A) in case they were Person C. The other participants will learn only at the end of the experiment which role they were assigned to: one of them Person B and the other one Person C. The earnings of all participants results from the assignment of roles and the decision taken by them.

The two other participants have to indicate how many reduction points they would assign you for each of your possible decisions (0 €, 2 €, or 5 €). Only the decision of Person C that corresponds to you actual decision will be implemented.

Punisher

One of the participants was assigned to the role of Person A. You and the remaining participants, who was not assigned to the role of Person A, will, at first, both take a decision in the role of Person C. You will both indicate how many reduction points you would assign to Person A in case you were Person C. You will both learn only at the end of the experiment which role you were assigned to: one of you Person B and the other one Person C. The earnings of all participants results from the assignment of roles and the decision taken by them.

You both have to indicate how many reduction points you would assign you for each of the possible decisions (0 €, 2 €, or 5 €) of Person A. Only the decision of Person C that corresponds to you actual decision will be implemented.

Before you take your decision on the next page, please answer the following two questions.

1. How much would Person A, B and C earn, if Person A gives 0 € to Person B and Person C has assigned 1 reduction point to Person A for that case?
2. How much would Person A, B and C earn, if Person A gives 5 € to Person B and Person C has assigned 0 reduction point to Person A for that case?

Instructions after games

SOCIAL

All participants have completed all tasks. Please stand up and wait until an assistant calls you cabin number. When you hear your cabin number, please read the following text verbatim.

Self-awareness questionnaire

Please indicate to which extent following statements describe yourself in this moment. Please answer based on how you feel now, in this moment – not based on how you feel generally.

Please answer on a scale from "does not apply at all" to "applies extremely well". There are no right or wrong answers. It is only important that you answer truthfully.

1. Right now, I am keenly aware of everything in my environment. (Surroundings)
2. Right now, I am conscious of my inner feelings. (Private)
3. Right now, I am concerned about the way I present myself. (Public)
4. Right now, I am self-conscious about the way I look. (Public)
5. Right now, I am conscious of what is going on around me. (Surroundings)
6. Right now, I am reflective about my life. (Private)
7. Right now, I am concerned about what other people think of me. (Public)
8. Right now, I am aware of my innermost thoughts. (Private)
9. Right now, I am conscious of all objects around me. (Surroundings)

Reputation questionnaire

The following question relate to the four tasks that you just completed.

Please think about how you felt during the tasks and indicate to which extent the following statements apply. Please answer on a scale from "I completely disagree" to "I completely agree".

1. During the task I did not think about what other participants would say about me.
2. It's important that the other participants will accept me.
3. During the task, I thought about how the other participants would think about me.
4. It's important to me that the other participants have a positive evaluation about me.

BIG-5 questionnaire

To which extent do the following statements apply to you?

Please indicate on the given scale which one corresponds your assessment! I see myself as someone who ...

1. ... is reserved
2. ... is generally trusting
3. ... tends to be lazy
4. ... is relaxed, handles stress well
5. ... has few artistic interests
6. ... is outgoing, sociable
7. ... tends to find fault with others
8. ... does a thorough job
9. ... gets nervous easily
10. ... has an active imagination

Additional questionnaire

1. How many studies at the Kölner Laboratorium für Wirschaftsforschung did you take part in (excluding this one)? (If you are not sure, guess.)
2. How much disposable income to you have per month (excluding rent)?
3. How well do you remember the online study?
4. What is more important for you – the rules and norms of society or your own rules and norms?
5. Would you feel ashamed to tell someone else that she broke a rule or a norm?
6. Would you be afraid to tell someone else that she broke a rule or a norm?
7. Would you describe yourself as a good or as a less good person?
8. Where would you locate yourself on the political spectrum from "very leftist" to "very rightist"?
9. How do you feel?
10. Did you enjoy this study?
11. You are welcome to leave a comment to the study.

C.2 Additional analyses

This Appendix contains further details, tables and graphs which complement our main analysis.

C.2.1 Attrition

As subjects participated in an online and a lab session that were about 4 weeks apart, we observe some attrition (24%). Here, we check whether attrition was systematic, as this might threaten the validity of our results. First, we check whether attrition is correlated with any of the observable characteristics elicited in the online study. Table C1 shows the results of a probit regression in which the dependent variable is a dummy equal to one if the subjects came to the lab and zero if the subject attrited. None of the observable characteristics predicts attrition.

	(1)
Female (=1)	0.028
	(0.048)
Siblings	-0.023
	(0.022)
Age	0.002
	(0.004)
Study (=1)	0.036
	(0.104)
Observations	330

Table C1: Probit model for attrition on observable characteristics

Note: Estimation of probit model with dummy variable for whether a subject participated also in the lab session or only in the online session as the dependent variable, and socio-demographic variables collected in the online session as independent variables. Coefficients represent average marginal effects. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Second, we go one step further and check whether the personal and social norm ratings differ between those who participated in the lab and the online session, and those who participated only in the online session. We compare the distribution of the two appropriateness ratings across the two samples for each action in the four games, for both personal and social norms. We run altogether 78 Chi-squared tests with simulated p-values over 10.000 replications, and use the Bonferroni correction to account for multiple hypotheses testing at the game level for personal and social norms separately. Only one out of the 78 tests turns out significant. Thus, the norm ratings across the two samples are highly consistent. Altogether, the observed attrition does not present a problem for the interpretation our results.

C.2.2 Estimation of the utility framework

To estimate our utility framework (Equation 3.1), we follow the common approach taken by the current literature and use a conditional (fixed-effects) logit choice model (see, e.g., Krupka and Weber, 2013; Gächter et al., 2013; Krupka et al., 2017). To estimate the model, we first reshape our dataset for each game.

For DG, we expand each individual observation to the amount of actions the subject in the role of Dictator could choose from (give €0, €1, ..., €10; 11 observations in total). We then generate a new dependant variable which equals one if the subject chose the given action and zero if she did not. We regress this outcome on characteristics of that potential action, which are the three dependent variables from our utility framework. The first variable is the monetary payoff. Here (as well as in the other games), we assume a linear restriction on the utility from money $V()$. Hence, in the DG, the monetary payoff is equal the amount of euros a subject would receive by choosing the particular action. The second dependent variable is the social norm appropriateness rating assigned by the subject to that action, while the third is the personal norm appropriateness rating she assigned to that action. The regression takes into account that each of the 11 observations stems from one subject's choice in the DG.

The same approach was taken for the other three games with the necessary adjustments. In the DGT, subjects had eight potential actions, which translates into eight observations per subject. In the UG, subjects playing as Receivers had eleven actions to choose from; hence, this translates into eleven observations per subject. To get the Receivers' monetary payoff in the UG, we calculated their expected payoff for each rejection threshold (i.e., each potential action) using the distribution of all proposers' offers. Finally, in the TPP, each subject playing as a Punisher made three choices, as she had to indicate her punishment choice for each potential action of the Dictator (strategy method). Each of these choices consisted of three potential actions; hence, we expanded the dataset to 9 observations per participant and the observations were grouped into three separate choices.[1]

[1]During the first day of data collection, subjects in the TPP game were exposed to a non-obstructive software issue. To avoid any potential bias in our estimation, we do not include the data from the TPP game collected during the first day in the analysis. To exclude that this affects our findings, we performed a robustness check by including this data. All reported results in the study stay robust to inclusion of this data.

C.2.3 Personal norms, social norms and behavior

We here report complementary information to our main results. Table C2 reports the estimation of our framework in the SOCIAL treatment. All coefficients on the personal norm ratings are significant across all games as well as in the pooled regression, confirming that Result 3.2 also holds in SOCIAL. The fact that personal norm coefficients are comparable with Table 3.1, and that social norm coefficients on average increase, reflects what we report in Result 3.3.

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.805***	0.146***	0.635***	0.676***	0.265***
	(0.118)	(0.039)	(0.176)	(0.138)	(0.031)
Social norm rating	2.149***	0.973***	0.922**	0.256	0.966***
	(0.381)	(0.223)	(0.363)	(0.240)	(0.134)
Personal norm rating	1.631***	0.691***	0.655*	1.076***	0.944***
	(0.335)	(0.206)	(0.340)	(0.253)	(0.130)
Observations	1,353	1,107	693	486	3,639

Table C2: Conditional logit estimation of choice determinants in SOCIAL treatment

Note: Estimation of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

In table C3, we report the results from the model comparison exercise in the main analysis leading up to Result 3.4. Note that we perform the same exercise also for all models estimated using the average social norm, as described below. Also in this case our conclusions remain unaffected.

	DG		DGt		UG		TPP		All games	
Main analysis	(1a) Sn	(1b) Sn + Pn	(2a) Sn	(2b) Sn + Pn	(3a) Sn	(3c) Sn + Pn	(4a) Sn	(4b) Sn + Pn	(5a) Sn	(5b) Sn + Pn
PRIVATE										
Log Likelihood	-240.9878	-229.6987	-238.1655	-231.4568	-119.6405	-116.6447	-116.2936	-110.7436	-738.6543	-707.9134
Likelihood test		22.58 (<0.001)		13.42 (<0.001)		5.99 (0.014)		11.10 (<0.001)		61.48 (<0.001)
Bayesian IC	496.4597	481.1237	490.4139	484.0379	252.3946	252.9598	245.0324	240.1549	1493.767	1440.514
SOCIAL										
Log Likelihood	-178.4372	-165.716	-230.9496	-225.3543	-109.3752	-107.5697	-126.0557	-116.3777	-686.1123	-659.378
Likelihood test		25.44 (<0.001)		11.19 (<0.001)		3.61 (0.057)		19.36 (<0.001)		53.47 (<0.001)
Bayesian IC	371.2947	353.0622	475.918	471.7368	231.8324	234.7624	264.4837	251.3141	1388.624	1343.354
Robustness check										
PRIVATE										
Log Likelihood	-226.5024	-219.6342	-234.6828	-227.035	-113.9186	-110.5374	-118.0499	-108.3958	-728.3834	-689.2193
Likelihood test		13.74 (<0.001)		15.30 (<0.001)		6.76 (0.009)		19.31 (<0.001)		68.82 (<0.001)
Bayesian IC	467.4889	457.1598	483.4483	475.1942	240.9508	240.7451	248.545	235.4592	1473.225	1412.636
SOCIAL										
Log Likelihood	-136.1462	-130.4522	-194.4412	-188.4233	-100.5577	-96.46517	-107.9209	-99.6109	-583.317	-558.7632
Likelihood test		11.39 (<0.001)		12.04 (<0.001)		8.19 (0.004)		16.62 (<0.001)		49.11 (<0.001)
Bayesian IC	286.7125	282.5347	402.9012	397.8748	214.1975	212.5534	228.2143	217.7805	1183.033	1142.125

Table C3: Model comparison

Note: Comparisons of log-likelihoods and Bayesian information criteria between models which include monetary payoff and social norms as predictors (Sn columns) and models which additionally include personal norms as a predictor (Sn + Pn columns). Comparisons are accompanied by likelihood ratio tests (p value in brackets) which are reported for the estimation of all individual games (Columns 1a - 4b) and all games together (5a - 5b), separately for PRIVATE and SOCIAL treatment. The upper panel reports comparisons following our main approach where we use the individual belief about social norms for the social norm rating, while the lower panel reports the comparisons for the robustness check where we use the average social norm for the social norm rating.

C.2.4 Robustness checks

As described in Section 3.3.3, we provide two robustness checks for our results. In the first, we want to rule out consistency as a potential explanation of our results. The regressions reported in Table C4 confirm Result 3.2. Personal norms remain a strong and stable predictor of behavior.

Regression (1) to (5) are performed with subjects who report a score below 6 when asked how well they remember the online experiment on a Likert scale from 1 to 7. Data are pooled across the PRIVATE and SOCIAL treatment to guarantee enough power. In regression (6), we only include subjects that score below the midpoint of our scale. We perform this regression by pooling all our games together and not for each game separately, as the number of observations decreases significantly.

	Memory < 6					Memory < 4
	DG	DGt	UG	TPP	All games	All games
	(1)	(2)	(3)	(4)	(5)	(6)
Monetary payoff	0.688***	0.226***	0.588***	0.799***	0.344***	0.343***
	(0.078)	(0.033)	(0.120)	(0.114)	(0.025)	(0.034)
Social norm rating	1.155***	0.746***	0.643**	0.390**	0.638***	0.333**
	(0.265)	(0.188)	(0.275)	(0.174)	(0.102)	(0.139)
Personal norm rating	1.462***	0.651***	0.961***	0.948***	0.987***	1.005***
	(0.241)	(0.160)	(0.264)	(0.185)	(0.098)	(0.135)
Observations	2,178	1,782	1,078	774	5,812	2,731

Table C4: Conditional logit estimation of choice determinants for robustness check of consistency

Note: Estimation of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as independent variables. The sample is restricted to subjects with a given score on the question of how well they remember the online session. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

In our second set of robustness checks, we use the average social norm rating for a given action instead of a subject's belief to re-run our complete analysis. In line with the literature using this approach (see., e.g., Krupka and Weber, 2013; Gächter and Schulz, 2016), we estimate a conditional logit choice model and calculate bootstrapped standard errors. More in details, as the average social norm ratings may suffer from a sampling error, we bootstrap 500 replications to calculate the errors. For each replication, we resample (with replacement) from the norm rating data to calculate the average of the social norm for that particular replication, and then resample (with replacement) from our behavioral data to conduct the replication. Table C5 displays the results of these regressions for the PRIVATE treatment. This confirms Result 3.2, namely that personal norms are a strong and stable predictor of behavior.

In Table C6, we provide a robustness check of Result 3.3. The interaction between average social norm ratings (as constructed for this robustness check) and the SOCIAL treatment are significant for the DG and DGT, as well as for all games pooled together. Also, the interaction between average social norm ratings and the PRIVATE treatment is insignificant in all regression models. Finally, we also report the estimations performed only for the SOCIAL treatment in Table C7. As expected, both personal and social norms ratings remain significant predictors of behavior, and the coefficients observed for personal norms remain comparable to those in the PRIVATE treatment.

	DG	DGt	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	1.179***	0.542	0.382	1.022***	0.520***
	(0.355)	(0.626)	(0.495)	(0.186)	(0.040)
Social norm rating (avg.)	2.649***	2.096	1.945	0.999***	1.101***
	(1.019)	(3.396)	(1.184)	(0.308)	(0.190)
Personal norm rating	0.986***	0.755***	0.721**	0.795***	0.880***
	(0.281)	(0.255)	(0.288)	(0.263)	(0.140)
Observations	1,397	1,143	704	504	3,748

Table C5: Conditional logit estimation of choice determinants in PRIVATE treatment with average social norm

Note: Estimation of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, average social appropriateness rating, and personal appropriateness rating of the action as independent variables. Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

	DG	DGt	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	1.357***	0.595	0.545	1.022***	0.512***
	(0.270)	(0.528)	(0.353)	(0.148)	(0.029)
Social norm rating (avg.)	3.161***	2.382	1.997*	0.999***	1.079***
	(0.799)	(2.863)	(1.029)	(0.323)	(0.190)
Personal norm rating	0.984***	0.757***	0.762***	0.795***	0.878***
	(0.286)	(0.257)	(0.289)	(0.260)	(0.138)
Social norm rating (avg.)	1.388***	2.141***	0.733	0.973*	1.832***
× SOCIAL	(0.414)	(0.501)	(0.836)	(0.540)	(0.313)
Personal norm rating	0.036	-0.102	-0.048	0.013	-0.067
× SOCIAL	(0.407)	(0.342)	(0.388)	(0.362)	(0.201)
Observations	2,750	2,250	1,397	990	7,387

Table C6: Conditional logit estimation of choice determinants interacted with SOCIAL treatment with average social norm

Note: Estimations of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action, as well as an interaction term between personal and average social norm ratings and the SOCIAL treatment as independent variables. Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

	DG	DGt	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	1.660***	0.682	1.432**	1.022***	0.500***
	(0.315)	(0.744)	(0.568)	(0.248)	(0.038)
Social norm rating (avg.)	5.352***	4.999	4.156***	1.972***	2.874***
	(0.926)	(3.915)	(1.294)	(0.444)	(0.259)
Personal norm rating	1.001***	0.661***	0.827***	0.808***	0.808***
	(0.303)	(0.238)	(0.319)	(0.245)	(0.148)
Observations	1,353	1,107	693	486	3,639

Table C7: Conditional logit estimation of choice determinants in Social treatment with average social norm

Note: Estimations of conditional logit choice model with dummy variable indicating whether the subjects chose the particular action as dependent variable, and monetary payoff, average social appropriateness rating, and personal appropriateness rating of the action as independent variables. Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

C.2.5 Further evidence on personal and social norms

	No. of subjects	Games rated
Group 1	67	Charitable giving game, Dictator game with entitlement, Lying game, Ultimatum game with computer first move
Group 2	46	Dictator game (DG), Charitable giving game with entitlement, Ultimatum game (UG), Trust game
Group 3	47	Dictator game with tax (DGt), Third party punishment game (TPP), Public good game, Vignettes

Table C8: Additional games.

The data for these additional experiments were collected during July and September 2017 at the Bonn Econ Lab. Subjects were divided in three groups and each group faced the norm elicitation task for a subset of the games (see Table C8) after an unrelated experiment. Subjects had to rate the personal and social appropriateness of each action available to the individual in the game or vignette presented to them. As in our main experiment, subjects were incentivized to guess the most common social appropriateness score of the given action in the session, while no incentives were provided for stating one's personal appropriateness score. All games and vignettes are described below.

Charitable giving game. An individual is given €10 and has to decide how much to give to a charity. She can give any integer amount between €0 and €10. The charity was UNICEF, an internationally renowned organization dedicated to provide humanitarian and developmental aid to children worldwide.

Charitable giving game with entitlement. Also here an individual has to decide how much out of €10 she wants to give to UNICEF. However, in this case, she has earned the €10 by answering to a questionnaire that lasted about 30 minutes.

Dictator game with entitlement. Similarly to the DG in the main analysis, an individual has €10 and can decide how much to give to another individual in the lab. Before this, however, both individuals have to work on a tiresome task for 20 minutes. They are given a series of matrices containing ones and zeros and have to count the number of zero in each matrix. The one who manages to complete more of such matrices is given the €10, the other €0.

Lying game. An individual is given a six-sided die and can privately roll it once. She gets the amount she report in euros. The participant rolls a one and can decide which number to report (from one to six).

Ultimatum game with computer first move. The structure of this game is analogous to that of the UG in our main experiment. An individual is given €10 and can offer any integer amount to another individual. If the individual accepts the offer, both get the proposed amounts. If she rejects it, they both earn nothing. However, the proposed amount is determined by a random device. The Responder has to state the minimum offer she is willing to accept.

Trust game. An individual receives €4 and a second one €0. The first individual can send any integer amount to the other one. This amount is tripled. The second individual can then decide how much she wants to send back to the first one. In the situation described the first individual sends €3 and the second participant had to decide how much of the €9 she received she wants to send back.

Public good game. Four individuals are grouped together and all receive €5. They, then, simultaneously decide how to allocate the €5 between a private and a common account. They can keep all the money in the private, while the money in the common account is summed up, multiplied by two and shared equally amongst all members.

Vignettes.

1. "Your neighbour pays a painter under the table and thus evades taxes."
2. "The chair of a commission at the university rejects a weak candidate to hire the daughter of a good friend."
3. "A woman who is moving out of her flat sells her couch which she bought for €1.500 for €2000."
4. "A freelancer eats at a restaurant with his friends for his birthday and deducts the check from his taxes."
5. "An employee of a firm calls in sick to prolong his holiday."
6. "A young man who finished university two years ago uses his old student card to drive on the train."
7. "A customer notices that he has been given €5 too much change at the supermarket, but keeps them."
8. "An acquaintance buys a highly polluting vintage car and drives it around just for fun in his free time."
9. "A colleague working from home claims to have worked for more hours than she actually did."
10. "An acquaintance who has purchased an insurance for his smart phone keeps it in the water to get a new one just before the insurance expires."

Table C9 displays the correlations and percent of non-zero differences for all seven additional games and the vignettes.

	correlation coefficient	% of non-zero differences		correlation coefficient	% of non-zero differences
Charitable giving game	0.53	66.35%	Ultimatum game with computer first move	0.56	58.62%
Charitable giving game with entitlement	0.54	61.41%	Trust game	0.53	76.74%
Dictator game with entitlement	0.53	62.42%	Public good game	0.75	47.87%
Lying game	0.65	50.25%	Vignettes	0.51	60.43%

Table C9: Correlation and % of non-zero differences in additional games.

Appendix D

D.1 Additional material

D.1.1 Instructions

Welcome!

Thank you for participating in this experiment. Please read the instructions below carefully. You will receive 2 € for your participation in today's experiment. You will have the possibility of earning more money during the experiment. The amount of these additional earnings depends on your own decisions, on the decisions of other participants, and on chance. Once the experiment is finished, you will also receive the payment from the online experiment. After the experiment, there will be a short questionnaire.

Please avoid any communication with your neighbors during the experiment. Switch off your mobile phone and remove everything you do not need for the experiment from the table. If you have any questions, please raise your hand and we will come to answer your questions at your seat.

Instructions

In this experiment, you will be matched with two other participants to form a group of three players. Each participant will be assigned a role within the group - a participant can be assigned either an active or a passive role. Two participants will be assigned an active role, Decision-Maker 1 and Decision-Maker 2. The third participant will be assigned the role of Passive Participant.

Decision-Maker 1 and Decision-Maker 2 will learn the cabin numer of the other decision-maker. They will not obtain any information about the cabin number of the Passive Participant. The Passive Participant will not be informed about the cabin number of the decision-makers.

The two decision-makers determine which of two options - Option 1 or Option 2 - they wish to choose. This choice influences how much money Decision-Maker 1, Decision-Maker 2, and the Passive Participant will earn. The amount of money the participants will earn with the chosen option also depends on which one of two events occurs: Event A or Event B. The participants have no influence on the occurrence of the these events

The computer determines which of the two events will occur at the beginning of the experiment for each group of three participants. The two events can occur with equal probability, i.e., the group can be in Event A with a probability of 50%, and in Event B with a probability of 50%.

Table 1 shows the different payments depending on the event occurred and the option chosen. The amounts are listed in the following order: Decision-Maker 1, Decision-Maker 2, and Passive

Participant.

	Event A	Event B
Option 1	6€, 6€, 0€	6€, 6€, 5€
Option 2	5€, 5€, 5€	5€, 5€, 5€

Table 1

- If the two decision-makers choose Option 1 and Event A occurs, Decision-Maker 1 and Decision-Maker 2 earn 6 € each, and the Passive Participant gets 0 €.
- If the two decision-makers choose Option 1 and Event B occurs, Decision-Maker 1 and Decision-Maker 2 earn 6 € each, and the Passive Participant gets 5 €.
- If the two decision-makers choose Option 2 and Event A occurs, Decision-Maker 1 and Decision-Maker 2 earn 5 € each, as does the Passive Participant.
- If the two decision-makers choose Option 1 and Event A occurs, Decision-Maker 1 and Decision-Maker 2 earn 5 € each, as does the Passive Participant.

The occurrence of Event A or Event B can be inferred from the content of an urn. The urn contains three red and two blue balls, if the group is in Event A. Or two red and three blue balls, if the group is in Event B. You can find a graphical visualization in Figure 1.

Figure 1

State A: 

State B: 

Before Decision-Maker 1 and Decision-Maker 2 choose between Option 1 and Option 2, they will obtain information about the content of the urn. However, they will never know the full content of the urn. Thus, they will never be certain about which of the two events occurred.

The computer will draw one ball from the urn and show it to both Decision-Maker 1 and Decision-Maker 2. (AsymmInfo: The computer will also draw another ball, without putting the first ball back in the urn, and it will show the second ball only to Decision-Maker 2. Decision-Maker 1 will never know which ball Decision-Maker 2 was shown.) (HighInfo: The computer will also draw another ball, without putting the first ball back in the urn, and it will again show the second ball to Decision-Maker 1 and Decision-Maker 2.)

In Event A, there is a higher probability that the computer will draw a red ball and a lower probability that it will draw a blue ball. Conversely, in Event B, there is a higher probability that the computer will draw a blue ball and a lower probability that it will draw a red ball.

After Decision-Maker 1 and Decision-Maker 2 have obtained this information, they can choose between the two options. It will be randomly determined which of the two cam make a first proposal. The selected decision-maker can propose either Option 1 or Option 2, or she can leave the decision to the other decision-maker. If she chooses the latter alternative, the other decision-maker will take a definitive choice for both decision-makers. In the other two cases, the proposal will be forwarded to the other decision-maker, who can either accept the proposal, with the consequence of the proposed option being implemented, or she can make another proposal. In the latter case, the new proposal will be forwarded to the first decision-maker. This procedure will go on until both decision-makers agree on one option. However, the decision-makers have up to five minutes time to reach an agreement. If no agreement is reached within the five minutes, both the two decision-makers and the third party will earn 0 €.

At the end of the experiment, there will be a short questionnaire. After you have filled in the questionnaire, we kindly ask you to remain seated. We will call the participants one by one, to ensure a confidential payment. Do you have any further questions? Then raise your hand and we will come to you. Before the start of the experiment, there will be some comprehension questions.

D.2 Additional analyses

Additional results supporting the main analysis are presented in this section.

D.2.1 Updating

Right after the main decision, all subjects answered incentivized questions aimed at testing their ability to form correct Bayesian posteriors about the state of the world they were in based on the draw from the urn. All subjects in LowInfo had to state the probability of State A (the bad state) after the draw of a red ball ($p(bad) = 0.6$). Subjects in AsymmInfo and HighInfo also had to answer one other question out of two possible ones. They had to either state the probability of State A after the draw of a red and a blue ball ($p(bad) = 0.5$), or after the draw of two red balls ($p(bad) = 0.75$). Subjects could enter any probability $0 \leq p \leq 1$ in intervals of 0.05 and received 1 € for each correct answer. Overall, 47% of subjects got the first answer exactly right and 63% updated in the correct direction, i.e., indicated a probability higher than 50%. 51% of subjects indicated a probability of exactly 0.5 in the second question, while 10% got the third question exactly correct and 45% updated in the right direction.

D.2.2 LowInfo treatment

In Section 4.3, I compared delegation choices between decision-makers in LowInfo and non-experts in AsymmInfo. Here, I compare their proposals, i.e., their behavior as first mover when they chose not to delegate. With a posterior of $p(bad) = 0.6$ the share of unethical proposals is 25% in LowInfo and 44% in AsymmInfo ($p = 0.420$). While with $p(bad) = 0.4$ it is 44% in LowInfo and 29% in AsymmInfo ($p = 0.314$). Overall, decision-makers in LowInfo and non-experts in AsymmInfo do not make different proposals indicating that the key component to look at is the behavior of experts.

D.2.3 Additional Results

An elicitation of decision-makers' feelings of responsibility towards the decision shows that decision-makers feel more responsible the more information they get, i.e., the more balls they draw. The reported feeling of responsibility is higher both in AsymmInfo compared to LowInfo (Wilcoxon rank-sum test $p = 0.03$) and in HighInfo compared to LowInfo ($p = 0.05$). In line with this finding, experts feel more responsible than non-experts in AsymmInfo ($p = 0.07$). Experts also feel overall more ashamed ($p = 0.07$) and guiltier ($p = 0.02$).

Bibliography

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for truth-telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Agerström, Jens, Rickard Carlsson, Linda Nicklasson, and Linda Guntell, "Using descriptive social norms to increase charitable giving: The power of local norms," *Journal of Economic Psychology*, 2016, *52*, 147–153.

Akerlof, George A, "The economics of caste and of the rat race and other woeful tales," *The Quarterly Journal of Economics*, 1976, pp. 599–617.

_ , "The market for "lemons": Quality uncertainty and the market mechanism," in "Uncertainty in economics," Elsevier, 1978, pp. 235–251.

_ , "A theory of social custom, of which unemployment may be one consequence," *The Quarterly Journal of Economics*, 1980, *94* (4), 749–775.

_ **and Dennis J Snower**, "Bread and bullets," *Journal of Economic Behavior & Organization*, 2016, *126*, 58–71.

_ **and William T Dickens**, "The economic consequences of cognitive dissonance," *The American Economic Review*, 1982, *72* (3), 307–319.

Allcott, Hunt and Todd Rogers, "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," *American Economic Review*, 2014, *104* (10), 3003–37.

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

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

_ **and Justin M Rao**, "The power of asking: How communication affects selfishness, empathy, and altruism," *Journal of Public Economics*, 2011, *95* (7-8), 513–520.

_ **and Ragan Petrie**, "Public goods experiments without confidentiality: a glimpse into fundraising," *Journal of public Economics*, 2004, *88* (7-8), 1605–1623.

Aquino, Karl and I.I. Reed, "The self-importance of moral identity.," *Journal of Personality and Social Psychology*, 2002, *83* (6), 1423.

Ariely, Dan, Anat Bracha, and Stephan Meier, "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially," *American Economic Review*, 2009, *99* (1), 544–55.

Balliet, Daniel, Craig Parks, and Jeff Joireman, "Social value orientation and cooperation in social dilemmas: A meta-analysis," *Group Processes & Intergroup Relations*, 2009, *12* (4), 533–547.

Bandura, Albert, Claudio Barbaranelli, Gian Vittorio Caprara, and Concetta Pastorelli, "Mechanisms of moral disengagement in the exercise of moral agency.," *Journal of Personality and Social Psychology*, 1996, *71* (2), 364.

Banerjee, Ritwik, "On the interpretation of bribery in a laboratory corruption game: moral frames and social norms," *Experimental Economics*, 2016, *19* (1), 240–267.

Barr, Abigail and Georgia Michailidou, "Complicity without connection or communication," *Journal of Economic Behavior & Organization*, 2017, *142*, 1–10.

Bartke, Simon, Andreas Friedl, Felix Gelhaar, and Laura Reh, "Social comparison nudges—Guessing the norm increases charitable giving," *Economics Letters*, 2017, *152*, 73–75.

Bartling, Björn and Urs Fischbacher, "Shifting the blame: On delegation and responsibility," *The Review of Economic Studies*, 2011, *79* (1), 67–87.

Battigalli, Pierpaolo and Martin Dufwenberg, "Guilt in games," *American Economic Review*, 2007, *97* (2), 170–176.

_ and _ , "Dynamic psychological games," *Journal of Economic Theory*, 2009, *144* (1), 1–35.

Bašić, Zvonimir, Armin Falk, and Simone Quercia, "Self-image, social image, and prosocial behavior," *mimeo*, 2020.

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

_ and _ , "Identity, morals, and taboos: Beliefs as assets," *The Quarterly Journal of Economics*, 2011, *126* (2), 805–855.

_ and _ , "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–64.

_ , Armin Falk, and Jean Tirole, "Narratives, Imperatives and Moral Reasoning," 2018.

Bergh, Donald D, David J Ketchen Jr, Ilaria Orlandi, Pursey PMAR Heugens, and Brian K Boyd, "Information asymmetry in management research: Past accomplishments and future opportunities," *Journal of Management*, 2019, *45* (1), 122–158.

Berkowitz, Leonard and Louise R Daniels, "Affecting the salience of the social responsibility norm: Effects of past help on the response to dependency relationships.," *The Journal of Abnormal and Social Psychology*, 1964, *68* (3), 275.

Beshears, John, James J Choi, David Laibson, Brigitte C Madrian, and Katherine L Milkman, "The effect of providing peer information on retirement savings decisions," *The Journal of Finance*, 2015, *70* (3), 1161–1201.

Bicchieri, Cristina, *The grammar of society: The nature and dynamics of social norms*, Cambridge University Press, 2005.

_ , "Norms, preferences, and conditional behavior," *Politics, Philosophy & Economics*, 2010, *9* (3), 297–313.

_ **and Hugo Mercier**, "Self-serving biases and public justifications in trust games," *Synthese*, 2013, *190* (5), 909–922.

Bó, Ernesto Dal and Pedro Dal Bó, ""Do the right thing:" The effects of moral suasion on cooperation," *Journal of Public Economics*, 2014, *117*, 28–38.

Böhm, Robert, Cornelia Betsch, and Lars Korn, "Selfish-rational non-vaccination: Experimental evidence from an interactive vaccination game," *Journal of Economic Behavior & Organization*, 2016, *131*, 183–195.

Bohnet, Iris, "The sound of silence in prisoner's dilemma and dictator games," in "Economics as a Science of Human Behaviour," Springer, 1999, pp. 177–194.

Bolton, Gary E and Axel Ockenfels, "ERC: A theory of equity, reciprocity, and competition," *American Economic Review*, 2000, *90* (1), 166–193.

Bolton, Patrick, Mathias Dewatripont et al., *Contract theory*, MIT press, 2005.

Bornstein, Gary and Ilan Yaniv, "Individual and group behavior in the ultimatum game: Are groups more "rational" players?," *Experimental Economics*, 1998, *1* (1), 101–108.

Bott, Kristina Maria, Alexander W Cappelen, Erik Sorensen, and Bertil Tungodden, "You've got mail: A randomised field experiment on tax evasion," 2017.

Brañas-Garza, Pablo, "Promoting helping behavior with framing in dictator games," *Journal of Economic Psychology*, 2007, *28* (4), 477–486.

Brodbeck, Felix C, Rudolf Kerschreiter, Andreas Mojzisch, and Stefan Schulz-Hardt, "Group decision making under conditions of distributed knowledge: The information asymmetries model.," *Academy of Management Review*, 2007, *32* (2), 459–479.

Bruner, Jerome, "The narrative construction of reality," *Critical Inquiry*, 1991, *18* (1), 1–21.

Burks, Stephen V and Erin L Krupka, "A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry," *Management Science*, 2012, *58* (1), 203–217.

Cappelen, Alexander W, Cornelius Cappelen, and Bertil Tungodden, "Second-best fairness under Limited information: The trade-off between false positives and false negatives," *NHH Dept. of Economics Discussion Paper*, 2018, (18).

_ , Karl O Moene, Erik Ø Sørensen, and Bertil Tungodden, "Needs versus entitlements—an international fairness experiment," *Journal of the European Economic Association*, 2013, *11* (3), 574–598.

_ , Trond Halvorsen, Erik Ø Sørensen, and Bertil Tungodden, "Face-saving or fair-minded: What motivates moral behavior?," *Journal of the European Economic Association*, 2017, *15* (3), 540–557.

Carlson, Ryan W, Michel Marechal, Bastiaan Oud, Ernst Fehr, and Molly Crockett, "Motivated misremembering: Selfish decisions are more generous in hindsight," 2018.

Chance, Zoë, Michael I Norton, Francesca Gino, and Dan Ariely, "Temporal view of the costs and benefits of self-deception," *Proceedings of the National Academy of Sciences*, 2011, *108* (Supplement 3), 15655–15659.

Charness, Gary and Martin Dufwenberg, "Promises and partnership," *Econometrica*, 2006, *74* (6), 1579–1601.

_ and Matthias Sutter, "Groups make better self-interested decisions," *Journal of Economic Perspectives*, 2012, *26* (3), 157–76.

Chytilova, Julie and Vaclav Korbel, "Individual and group cheating behavior: a field experiment with adolescents," Technical Report, IES Working Paper 2014.

Cialdini, Robert B, Carl A Kallgren, and Raymond R Reno, "A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior," in "Advances in experimental social psychology," Vol. 24, Elsevier, 1991, pp. 201–234.

_ , Linda J Demaine, Brad J Sagarin, Daniel W Barrett, Kelton Rhoads, and Patricia L Winter, "Managing social norms for persuasive impact," *Social influence*, 2006, *1* (1), 3–15.

_ , Raymond R Reno, and Carl A Kallgren, "A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places.," *Journal of personality and social psychology*, 1990, *58* (6), 1015.

Conrads, Julian, Bernd Irlenbusch, Rainer Michael Rilke, and Gari Walkowitz, "Lying and team incentives," *Journal of Economic Psychology*, 2013, *34*, 1–7.

Cox, James C, "Trust, reciprocity, and other-regarding preferences: Groups vs. individuals and males vs. females," in "Experimental business research," Springer, 2002, pp. 331–350.

Croson, Rachel and Melanie Marks, "The effect of recommended contributions in the voluntary provision of public goods," *Economic Inquiry*, 2001, *39* (2), 238–249.

Dana, Jason, George Loewenstein, Roberto A Weber, D De Cremer, and AE Tenbrunsel, "Ethical immunity: How people violate their own moral standards without feeling they are doing so," *Behavioral business ethics: Shaping an emerging field*, 2012, pp. 201–219.

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

Ditto, Peter H, David A Pizarro, and David Tannenbaum, "Motivated moral reasoning," *Psychology of Learning and Motivation*, 2009, *50*, 307–338.

Dreber, Anna, Tore Ellingsen, Magnus Johannesson, and David G Rand, "Do people care about social context? Framing effects in dictator games," *Experimental Economics*, 2013, *16* (3), 349–371.

Dufwenberg, Martin and Georg Kirchsteiger, "A theory of sequential reciprocity," *Games and Economic Behavior*, 2004, *47* (2), 268–298.

Ellman, Matthew and Paul Pezanis-Christou, "Organizational structure, communication, and group ethics," *American Economic Review*, 2010, *100* (5), 2478–91.

Elster, Jon, "Social norms and economic theory," *Journal of Economic Perspectives*, 1989, *3* (4), 99–117.

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

Epley, Nicholas and Thomas Gilovich, "The mechanics of motivated reasoning," *Journal of Economic Perspectives*, 2016, *30* (3), 133–40.

Erat, Sanjiv, "Avoiding lying: The case of delegated deception," *Journal of Economic Behavior & Organization*, 2013, *93*, 273–278.

Ewers, Mara and Florian Zimmermann, "Image and misreporting," *Journal of the European Economic Association*, 2015, *13* (2), 363–380.

Exley, Christine L, "Excusing selfishness in charitable giving: The role of risk," *The Review of Economic Studies*, 2015, *83* (2), 587–628.

Falk, Armin, "Facing Yourself-A Note on Self-image," 2017.

_ and Florian Zimmermann, "Information processing and commitment," *The Economic Journal*, 2018, *128* (613), 1983–2002.

_ and Nora Szech, "Organizations, diffused pivotality and immoral outcomes," 2013.

_ and Urs Fischbacher, "A theory of reciprocity," *Games and Economic Behavior*, 2006, *54* (2), 293–315.

_ , **Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde**, "Global evidence on economic preferences," *The Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.

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

_ **and Simon Gächter**, "Fairness and retaliation: The economics of reciprocity," *Journal of Economic Perspectives*, 2000, *14* (3), 159–181.

_ **and Urs Fischbacher**, "Third-party punishment and social norms," *Evolution and Human Behavior*, 2004, *25* (2), 63–87.

_ **, Erich Kirchler, Andreas Weichbold, and Simon Gächter**, "When social norms overpower competition: Gift exchange in experimental labor markets," *Journal of Labor economics*, 1998, *16* (2), 324–351.

Feiler, Lauren, "Testing models of information avoidance with binary choice dictator games," *Journal of Economic Psychology*, 2014, *45*, 253–267.

Ferraro, Paul J and Michael K Price, "Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment," *Review of Economics and Statistics*, 2013, *95* (1), 64–73.

Festinger, Leon, "A theory of social comparison processes," *Human Relations*, 1954, *7* (2), 117–140.

_ , *A theory of cognitive dissonance*, Vol. 2, Stanford university press, 1962.

Fiedler, Susann, Andreas Glöckner, Andreas Nicklisch, and Stephan Dickert, "Social value orientation and information search in social dilemmas: An eye-tracking analysis," *Organizational Behavior and Human Decision Processes*, 2013, *120* (2), 272–284.

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

_ **and Franziska Föllmi-Heusi**, "Lies in disguise—an experimental study on cheating," *Journal of the European Economic Association*, 2013, *11* (3), 525–547.

Foerster, Manuel and Joel J van der Weele, "Denial and Alarmism in Collective Action Problems," 2018.

_ **and** _ , "Persuasion, justification and the communication of social impact," 2018.

Forsythe, Robert, Joel L Horowitz, Nathan E Savin, and Martin Sefton, "Fairness in simple bargaining experiments," *Games and Economic Behavior*, 1994, *6* (3), 347–369.

Frey, Bruno S and Stephan Meier, "Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment," *American Economic Review*, 2004, *94* (5), 1717–1722.

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

_ , Daniele Nosenzo, and Martin Sefton, "Peer effects in pro-social behavior: Social norms or social preferences?," *Journal of the European Economic Association*, 2013, *11* (3), 548–573.

Galbiati, Roberto and Pietro Vertova, "Obligations and cooperative behaviour in public good games," *Games and Economic Behavior*, 2008, *64* (1), 146–170.

Geanakoplos, John, David Pearce, and Ennio Stacchetti, "Psychological games and sequential rationality," *Games and Economic Behavior*, 1989, *1* (1), 60–79.

Ging-Jehli, Nadja R, Florian Schneider, and Roberto A Weber, "On self-serving strategic beliefs," *University of Zurich, Department of Economics, Working Paper*, 2019, (315).

Gino, Francesca, Michael I Norton, and Roberto A Weber, "Motivated Bayesians: Feeling moral while acting egoistically," *Journal of Economic Perspectives*, 2016, *30* (3), 189–212.

_ , Shahar Ayal, and Dan Ariely, "Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel," *Psychological Science*, 2009, *20* (3), 393–398.

_ , _ , and _ , "Self-serving altruism? The lure of unethical actions that benefit others," *Journal of Economic Behavior & Organization*, 2013, *93*, 285–292.

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

Goeree, Jacob K, Charles A Holt, and Susan K Laury, "Private costs and public benefits: unraveling the effects of altruism and noisy behavior," *Journal of Public Economics*, 2002, *83* (2), 255–276.

Gollwitzer, M., K. Schmidthals, and C. Pöhlmann, "Relationalitäts-Kontextabhängigkeits-Skala (RKS): Entwicklung und erste Ansätze zur Validierung. (Berichte aus der Arbeitsgruppe "Verantwortung, Gerechtigkeit, Moral" Nr. 161)," *Trier: Universität Trier*, 2006.

Golman, Russell, George Loewenstein, Karl Ove Moene, and Luca Zarri, "The preference for belief consonance," *Journal of Economic Perspectives*, 2016, *30* (3), 165–88.

Govern, John M and Lisa A Marsch, "Development and validation of the situational self-awareness scale," *Consciousness and Cognition*, 2001, *10* (3), 366–378.

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

Grossman, Zachary and Joël J Van Der Weele, "Self-image and willful ignorance in social decisions," *Journal of the European Economic Association*, 2017, *15* (1), 173–217.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze, "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, 1982, *3* (4), 367–388.

Haisley, Emily C and Roberto A Weber, "Self-serving interpretations of ambiguity in other-regarding behavior," *Games and Economic Behavior*, 2010, *68* (2), 614–625.

Hamman, John R, George Loewenstein, and Roberto A Weber, "Self-interest through delegation: An additional rationale for the principal-agent relationship," *American Economic Review*, 2010, *100* (4), 1826–1846.

Iriberri, Nagore and Pedro Rey-Biel, "The role of role uncertainty in modified dictator games," *Experimental Economics*, 2011, *14* (2), 160–180.

Jackson, Matthew O and Massimo Morelli, "The reasons for wars: an updated survey," *The Handbook on the Political Economy of War*, 2011, *34*.

Kahan, Dan M, Ellen Peters, Maggie Wittlin, Paul Slovic, Lisa Larrimore Ouellette, Donald Braman, and Gregory Mandel, "The polarizing impact of science literacy and numeracy on perceived climate change risks," *Nature Climate Change*, 2012, *2* (10), 732.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler, "Fairness and the assumptions of economics," *Journal of Business*, 1986, pp. S285–S300.

Karlsson, Niklas, George Loewenstein, Jane McCafferty et al., "The economics of meaning," *Nordic Journal of Political Economy*, 2004, *30* (1), 61–75.

Kessler, Judd B and Stephen Leider, "Norms and contracting," *Management Science*, 2012, *58* (1), 62–77.

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

Kocher, Martin G, Simeon Schudy, and Lisa Spantig, "I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups," *Management Science*, 2017, *64* (9), 3995–4008.

Konow, James, "Fair shares: Accountability and cognitive dissonance in allocation decisions," *American Economic Review*, 2000, *90* (4), 1072–1091.

Krämer, Florentin, Klaus M Schmidt, Martin Spann, and Lucas Stich, "Delegating pricing power to customers: Pay what you want or name your own price?," *Journal of Economic Behavior & Organization*, 2017, *136*, 125–140.

Krupka, Erin and Roberto A Weber, "The focusing and informational effects of norms on pro-social behavior," *Journal of Economic Psychology*, 2009, *30* (3), 307–320.

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

_ , **Stephen Leider, and Ming Jiang**, "A meeting of the minds: Informal agreements and social norms," *Management Science*, 2017, *63* (6), 1708–1729.

Kugler, Tamar, Gary Bornstein, Martin G Kocher, and Matthias Sutter, "Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy," *Journal of Economic Psychology*, 2007, *28* (6), 646–657.

Lange, Paul AM Van, Wim BG Liebrand, and D Michael Kuhlman, "Causal attribution of choice behavior in three N-person prisoner's dilemmas," *Journal of Experimental Social Psychology*, 1990, *26* (1), 34–48.

Larson, Tara and C Monica Capra, "Exploiting moral wiggle room: Illusory preference for fairness? A comment," *Judgment and Decision Making*, 2009, *4* (6), 467.

Lazear, Edward P, Ulrike Malmendier, and Roberto A Weber, "Sorting in experiments with application to social preferences," *American Economic Journal: Applied Economics*, 2012, *4* (1), 136–63.

Liebrand, Wim BG, Ronald WTL Jansen, Victor M Rijken, and Cor JM Suhre, "Might over morality: Social values and the perception of other players in experimental games," *Journal of Experimental Social Psychology*, 1986, *22* (3), 203–215.

Lindbeck, Assar, "Incentives and social norms in household behavior," *The American Economic Review*, 1997, *87* (2), 370–377.

_ , **Sten Nyberg, and Jörgen W Weibull**, "Social norms and economic incentives in the welfare state," *The Quarterly Journal of Economics*, 1999, *114* (1), 1–35.

Luhan, Wolfgang J, Martin G Kocher, and Matthias Sutter, "Group polarization in the team dictator game reconsidered," *Experimental Economics*, 2009, *12* (1), 26–41.

Matthey, Astrid and Tobias Regner, "Do I really want to know? A cognitive dissonance-based explanation of other-regarding behavior," *Games*, 2011, *2* (1), 114–135.

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

McAdams, Dan P, *Power, intimacy, and the life story: Personological inquiries into identity*, Guilford Press, 1988.

McFadden, Daniel et al., "Conditional logit analysis of qualitative choice behavior," 1973.

Miller, Dale T and Michael Ross, "Self-serving biases in the attribution of causality: Fact or fiction?," *Psychological Bulletin*, 1975, *82* (2), 213.

Mohlin, Erik and Magnus Johannesson, "Communication: Content or relationship?," *Journal of Economic Behavior & Organization*, 2008, *65* (3-4), 409–419.

Moore, Celia, James R Detert, Linda Klebe Treviño, Vicki L Baker, and David M Mayer, "Why employees do bad things: Moral disengagement and unethical organizational behavior," *Personnel Psychology*, 2012, *65* (1), 1–48.

Morton, Rebecca B and Jean-Robert Tyran, "Let the experts decide? Asymmetric information, abstention, and coordination in standing committees," *Games and Economic Behavior*, 2011, *72* (2), 485–509.

Murphy, Ryan, Kurt Ackermann, and Michel Handgraaf, "Measuring social value orientation," *Judgment and Decision Making*, 2011, *6* (8), 771–781.

Offerman, Theo, Joep Sonnemans, and Arthur Schram, "Value orientations, expectations and voluntary contributions in public goods," *The Economic Journal*, 1996, pp. 817–845.

Ostrom, Elinor, "Collective action and the evolution of social norms," *Journal of Economic Perspectives*, 2000, *14* (3), 137–158.

Rabin, Matthew, "Incorporating fairness into game theory and economics," *The American Economic Review*, 1993, pp. 1281–1302.

Rammstedt, Beatrice and Oliver P John, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, 2007, *41* (1), 203–212.

Rodriguez-Lara, Ismael and Luis Moreno-Garrido, "Self-interest and fairness: Self-serving choices of justice principles," *Experimental Economics*, 2012, *15* (1), 158–175.

Romano, Angelo and Daniel Balliet, "Reciprocity outperforms conformity to promote cooperation," *Psychological Science*, 2017, *28* (10), 1490–1502.

Rutkowski, Gregory K, Charles L Gruder, and Daniel Romer, "Group cohesiveness, social norms, and bystander intervention.," *Journal of Personality and Social Psychology*, 1983, *44* (3), 545.

Saucet, Charlotte and Marie Claire Villeval, "Motivated Memory in Dictator Games," 2018.

Schwartz, Shalom H, "Normative explanations of helping behavior: A critique, proposal, and empirical test," *Journal of Experimental Social Psychology*, 1973, *9* (4), 349–364.

_ , "Normative influences on altruism," *Advances in Experimental Social Psychology*, 1977, *10* (1), 221–279.

_ **and John A Fleishman**, "Personal norms and the mediation of legitimacy effects on helping," *Social Psychology*, 1978, pp. 306–315.

Selten, Reinhard, "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes," in "in" Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie 1965.

Shafir, Eldar, Itamar Simonson, and Amos Tversky, "Reason-based choice," *Cambridge series on judgment and decision making. Research on judgment and decision making: Currents, connections, and controversies*, 1997, pp. 69–94.

Shalvi, Shaul, Francesca Gino, Rachel Barkan, and Shahar Ayal, "Self-serving justifications: Doing wrong and feeling moral," *Current Directions in Psychological Science*, 2015, *24* (2), 125–130.

_ , Jason Dana, Michel JJ Handgraaf, and Carsten KW De Dreu, "Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior," *Organizational Behavior and Human Decision Processes*, 2011, *115* (2), 181–190.

_ , Michel JJ Handgraaf, and Carsten KW De Dreu, "Ethical manoeuvring: Why people avoid both major and minor lies," *British Journal of Management*, 2011, *22*, S16–S27.

Shang, Jen and Rachel Croson, "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods," *The Economic Journal*, 2009, *119* (540), 1422–1439.

Shiller, Robert J, "Narrative economics," *American Economic Review*, 2017, *107* (4), 967–1004.

Sohrab, Serena G, Mary J Waller, and Seth Kaplan, "Exploring the hidden-profile paradigm: A literature review and analysis," *Small Group Research*, 2015, *46* (5), 489–535.

Soraperra, Ivan, Anton Suvorov, Jeroen Van de Ven, and Marie Claire Villeval, "Doing Bad to Look Good," *Revue Économique*, 2019, *70* (6), 945–966.

_ , Ori Weisel, Sys Kochavi, Margarita Leib, Hadar Shalev, Shaul Shalvi et al., "The bad consequences of teamwork," *Economics Letters*, 2017, *160*, 12–15.

Spranca, Mark, Elisa Minsk, and Jonathan Baron, "Omission and commission in judgment and choice," *Journal of Experimental Social Psychology*, 1991, *27* (1), 76–105.

Stasser, Garold and William Titus, "Pooling of unshared information in group decision making: Biased information sampling during discussion.," *Journal of Personality and Social Psychology*, 1985, *48* (6), 1467.

_ and _ , "Hidden profiles: A brief history," *Psychological Inquiry*, 2003, *14* (3-4), 304–313.

Suls, Jerry and Ladd Wheeler, *Handbook of social comparison: Theory and research*, Springer Science & Business Media, 2013.

Sutter, Matthias, "Deception through telling the truth?! Experimental evidence from individuals and teams," *The Economic Journal*, 2009, *119* (534), 47–60.

_ and Martin G Kocher, "Trust and trustworthiness across different age groups," *Games and Economic Behavior*, 2007, *59* (2), 364–382.

Tella, Rafael Di, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman, "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism," *American Economic Review*, 2015, *105* (11), 3416–42.

Tesser, Abraham, "Toward a self-evaluation maintenance model of social behavior.," 1985.

Treviño, Linda Klebe, Niki A Den Nieuwenboer, and Jennifer J Kish-Gephart, "(Un) ethical behavior in organizations," *Annual Review of Psychology*, 2014, *65.*

van der Weele, Joël J, Julija Kulisa, Michael Kosfeld, and Guido Friebel, "Resisting moral wiggle room: how robust is reciprocal behavior?," *American Economic Journal: Microeconomics*, 2014, *6* (3), 256–64.

Weisel, Ori and Shaul Shalvi, "The collaborative roots of corruption," *Proceedings of the National Academy of Sciences*, 2015, *112* (34), 10651–10656.

_ et al., "Social motives in intergroup conflict: Group identity and perceived target of threat," *European Economic Review*, 2016, *90*, 122–133.

Wiltermuth, Scott S, "Cheating more when the spoils are split," *Organizational Behavior and Human Decision Processes*, 2011, *115* (2), 157–168.

Wittenbaum, Gwen M, Andrea B Hollingshead, and Isabel C Botero, "From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm," *Communication Monographs*, 2004, *71* (3), 286–310.

Xiao, Erte, "Justification and conformity," *Journal of Economic Behavior & Organization*, 2017, *136*, 15–28.

Web References

BBC (2015). Volkswagen staff acted criminally, says board member. [online] Available at: https://www.bbc.com/news/business-34397426 [Accessed 22 Oct. 2019].

O'Kane, S. (2015). Volkswagen America's CEO blames software engineers for emissions cheating scandal. The Verge. [online] Available at: https://www.theverge.com/2015/10/8/9481651/volkswagen-congressional-hearing-diesel-scandal-fault [Accessed 22 Oct. 2019].

Eugenio Verrina

Kurt-Schumacher-Str. 10
D-53113 Bonn
Germany

Phone: (0049) 228 91416-152
verrina@coll.mpg.de
WEB: https://sites.google.com/view/eugenio-verrina/home

Current Position

Ph.D. candidate in Economics, University of Cologne
(Supervisors: Prof. Dr. Bettina Rockenbach and Prof. Dr. Engel)

Research Fellow, Max Planck Institute for Research on Collective Goods (Bonn)

Research Assistant, Chair for Experimental and Behavioral Economics

Memeber of C-SEB (Center for Social and Economic Behavior)

Research Interests

Behavioral and Experimental Economics: motivated beliefs, self and social image concerns and norms

Education

Fall 2018 Visiting PhD Student at University of Zurich, Department of Economics (Host: Prof. Roberto Weber)

2016- Ph.D. in Economics, MPI for Research on Collective Goods and University of Cologne

2014-2016 M.Sc. in Economics, University of Trento, (Supervisor: Prof. Matteo Ploner)

Spring 2016 Visiting period for research thesis, University of Vienna, Department of Applied Psychology

2011-2014 Double Bachelor Degree in Economics and Management, University of Trento and Dresden University of Technology

2011 Double High School Diploma (Italian and German), Deutsche Schule Genua

Working Papers

Hillenbrand, Adrian, and Eugenio Verrina. The differential effect of narratives on prosocial behavior MPI Collective Goods Discussion Paper 2018/16

Luigi Mittone, Matteo Ploner, and Eugenio Verrina, When the State Doesn't Play Dice: An Experimental Analysis of Cunning Fiscal Policies and Tax Compliance, CEEL Working Paper 2-17

Work in Progress (selected)

The Dark Side of Experts: Ethical Decision-making under Asymmetric Information in Teams

Social norms, personal norms and image concerns, with Zvonimir Bašić

Conferences & Talks

2020 Erasmus University Rotterdam; ZEW - Leibniz Center for European Economic Research;

2019 University of Innsbruck (2); THEEM 10th Thurgau Experimental Economics Meeting, University of Konstanz; IMEBESS 6th International Meeting on Experimental and Behavioral Social Sciences, Utrecht University; ESA European Meeting (Dijon); University of Genova;

2018 ESA World Meeting (Berlin); IMEBESS European University Institute (Florence);

2017 "The Shadow Economy, Tax Evasion and Informal Labor" 5th edition, University of Warsaw;

2016 Department of Applied Psychology, University of Vienna; Generations of Experiments, University of Trento

Summer Schools, Courses & co.

2019 DGPE PhD course: Psychological Game Theory, University of Copenhagen; WESSI: Winter Experimental Social Sciences Institute, NYU Florence;

2018 IMPRS Uncertainty Topics Workshop, LUISS Rome; Summer School "Behavioural Game Theory", University of East Anglia

2017 IMPRS Uncertainty Topics Workshop, University of Trento; 11th IMPRS Uncertainty Summer School, Friedrich Schiller University (Jena)

2016 IMPRS Uncertainty Topics Workshop, Friedrich Schiller University (Jena); 10th IMPRS Uncertainty Summer School, Friedrich Schiller University (Jena); Organiser and moderator of the public debate Neoclassical vs. Behavioural Economics, Rethinking Economics Trento;

2015 Experiment a BIT, "Euregio" meeting

Grants & Awards

2016 Prize, "2016 Merit Prize", University of Trento, Trento, Italy
Scholarship, Opportunities of international mobility for thesis research/final examination abroad, University of Trento, Italy

2015 Prize, "2015 Merit Prize", University of Trento, Trento, Italy

2013-2104 Scholarship, Double Degree Progam, University of Trento, Italy

Refereeing

Management Science; International Tax and Public Finance; Rationality & Society

Teaching

Summer term 2018 & 2019 and winter term 2019-20: Experimental Methods, Tutorials (Master level, University of Cologne)

Languages

English: Professional knowledge

German: Professional knowledge

French: Fluent knowledge

Italian: Mother tongue

Programming Languages and Scientific Applications

Python, Stata, R, LaTeX, z-Tree.

References

Bettina Rockenbach

University of Cologne
Chair for Experimental and Behavioral Economics
Email: bettina.rockenbach(at)uni-koeln.de
Phone: +49 221 470 8664

Cristoph Engel

Director of the Max Planck Institute
for Research on Collective Goods (Bonn)
Email: engel(at)coll.mpg.de
Phone: +49 228 91416-210

Matthias Sutter

Director of the Max Planck Institute
for Research on Collective Goods (Bonn)
Email: sutter(at)coll.mpg.de
Phone: +49 228 914 16 865

Roberto Weber

University of Zurich
Department of Economics
Email: roberto.weber(at)econ.uzh.ch
Phone: +41 44 634 36 88

Last updated: July 28, 2020

“Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Disseration selbstaändig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe del Quelle gekennzeichnet. Beu del Auswahl und Auswertung folgenden Materials Haben mir die nachstehend aufgefürten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich (zutreffendes bitte unterstreichen) geholfen:

Weitere Personen, neben den in der Einleitung der Dissertation aufgeführten Koautorinnen und Koautoren, waren an der inhaltlish-materiellen Erstellung del vorliegenden Dissertation nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Dissertation wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelect. Ich versichere, dass ich nacch bestem Wissem die reine Wahrheit gesagt und nichts verschwiegen habe.”

Eugenio Verrina