Normative Perception of Power Abuse*

Leonard Hoeft[†] **Wladislaw Mill**[‡] **Alexander Vostroknutov**[§¶]

October 2018

Abstract

We study how the powerful perceive power abuse, and how negative experience related to it influences the appropriateness judgements of the powerless. We create an environment conducive to unfair exploitation in a repeated Public Goods game where one player (punisher) is given a further ability to costlessly subtract money from others (victims). We find that punishers, who choose to abuse their power, rationalize their behavior by believing that free-riding, while forcing others to contribute to the public good, is not inappropriate. Victims of such abuse also start to believe that punishers' free-riding and punishment are justifiable. Both punishers and victims are not aware that their beliefs are adjusting in this way. In addition, subjects assigned to the role of power, regardless of how they use it, think that outside observers share their beliefs about the appropriateness of their actions. All these observations are explained by the Belief in a Just World Hypothesis, which states that people rationalize any wrongful acts in order to maintain a coherent picture of the world that is orderly and lawful. Our findings demonstrate the fearsome capacity of humans to exculpate abusive behavior by themselves and others.

JEL classifications: C91, C92
Keywords: power abuse, norms, belief in a just world, public goods, punishment.

*We would like to thank the participants of the Behavioral and Experimental Economics Network seminar in Rome (Sep 2018) for invaluable comments. All mistakes are ours.

[†]Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Straße 10, 53113 Bonn Germany. e-mail: hoeft@coll.mpg.de

[‡]Department of Economics, University of Mannheim, L7 3-5, 68131 Mannheim, Germany. e-mail: mill@uni-mannheim.de

[§]Center for Mind/Brain Sciences, University of Trento, Via delle Regole 101, 38123 Mattarello (TN), Italy. e-mail: a.vostroknutov@unitn.it

[¶]Corresponding author.

1 Introduction

Experimental economics has a long history of investigating prosocial behavior. The consensus is that, contrary to the predictions of models with selfish preferences, people act largely prosocially (Schroeder and Graziano, 2015), which is backed by a fairly uncontroversial norms proscribing selfishness (Cubitt *et al.*, 2011; Krupka and Weber, 2013). As Fehr and Schurtenberger (2018) discuss at length, the normative foundations seem to permeate most social interactions. Nevertheless, unjust conditions and behavior are pervasive and hard to eradicate. Even developed countries with functioning legal and social systems witness high inequality and unfair distribution of power (Acemoglu *et al.*, 2015; Rose-Ackerman and Palifka, 2016). Indeed, much of the policy debate involves arguing that some part of society is disproportionately favored and, thereby, fails to contribute to the community: they, essentially, "play a rigged game" (Acemoglu and Robinson, 2008; Dal Bó *et al.*, 2009). This is true despite the fact that most modern societies feature the institutions that lead to prosocial behavior on interpersonal level.

The mismatch between individual prosociality and corruption among the powerful may originate from the differences in the normative perception of wrongful acts. On the one hand, the direct and forceful subjugation or mistreatment of others is considered extremely morally inappropriate, which results in few such problematic behaviors in modern societies. Even authoritarian states avoid clear moral violations and choose to veil subjugation of their subjects behind normative reasons (Beetham, 2013). On the societal level, only a small percentage of the population openly violate fundamental norms of fairness and respect for basic human rights in direct interactions with others. Those who steal or harm others are quickly ostracized and often considered to be antisocial or dangerous.

However, on the other hand, institutions that promote public welfare regularly create unfair opportunities for their functionaries at the expense of the general population. The rich and the powerful are able to exploit their privileged position in questionable ways. Such behaviors often take form of hypocritical enforcement of institutional rules that the enforcers do not adhere to themselves. Examples include police officers using illegal violence, politicians using their influence to attain atypical benefits, doctors using their connections for special treatment, and managers forcing their coworkers to invest in shared projects that they themselves skimp on. The ubiquity of this kind of practices and the ostensible perception that they are more benign than direct harm may be explained by their indirect consequences and the dilution of norms determining appropriate behavior in complex institutions. Moreover, it is easy to make excuses on the grounds that, even though an individual with power might use his position for personal

benefit, he still provides an important social service. In support of these views of normative perception of power abuse, previous research has shown that in complex environments moral disagreement is pervasive (Reuben and Riedl, 2013); people are reluctant to harm others in a personal and direct way, while harming them as a side effect seems more permissible (Greene *et al.*, 2009); there is a tendency to justify one's questionable actions with self-serving beliefs about the behavior of others (Di Tella *et al.*, 2015). There is also a substantial evidence that people victimized by an unfair treatment may react adversely to it: experiencing unfair behavior makes the punishment of similar acts in the future less likely (Herz and Taubinsky, 2017); witnessing social norm violations leads to less trust (Banerjee, 2016); experimental subjects from countries with a high corruption index are more likely to lie (Gächter and Schulz, 2016). The reaction to observed norm violations can be "contagious": criminal behavior is often spatially correlated (Glaeser *et al.*, 1996; Zenou, 2003), which, according to the proponents of the "broken windows hypothesis," is due to norm violations signaling a lack of commitment of a society to follow norms (Wilson and Kelling, 1982). Similarly, Fisman and Miguel (2007) observed that diplomats from corrupt countries committed more parking violations. On the institutional level, Tabellini (2008, 2010) shows that normative values in the regions, which experienced the rule of despotic institutions in the past, are less likely to be consistent with "generalized morality," or the norms of good conduct, than those in the countries that did not endure such rule.[1]

In spite of all this evidence that shows how detrimental corruption can be, the relationship between abuse of institutional power and associated normative perceptions of it remains, in large parts, unclear. Does everybody agree on the norms regarding indirect harm and abuse of power? Do the abusers simply use their advantageous position out of selfishness, or rationalize their behavior? Why is bad behavior contagious? Do people simply imitate others to not be taken advantage of, or do they also assimilate bad norms imposed by abusive institutions?

In this study we tackle these questions by experimentally investigating abuse of power and its causal relationships with normative perceptions by various parties. We implement a Public Goods game that allows one powerful participant (punisher), who fulfills the role of a sanctioning authority, to dictate contribution norms, while being free to exempt himself from them (Hoeft and Mill, 2017). Unlike the established designs, where all players have the means to punish others (e.g., Fehr and Gächter, 2000), this game models the ambivalence of indirect abuse of power: not contributing while forcing others to do so is unfair, but enforcing high contribution norms is beneficial, even if the punisher does not comply himself. In this setting, we investigate the motives of power abusers, the effect that experience of abuse has on the perceived appropriateness of punisher's actions, and the normative perceptions of outsiders who do not play the Public Goods game. Specifically, in order to understand how the powerful, the powerless,

[1] See also Becker *et al.* (2015). It should be mentioned that the reverse process has also been documented: Lowes *et al.* (2017) report the results of a field experiment showing that strong institutions in the past crowd out rule-following behavior today.

and uninvolved third parties perceive power abuse, or its absence, we elicit their beliefs about the prevalent norms of behavior *in their own reference group*. This allows us to see if there are differences in normative perceptions of the same situation generated by either being assigned to the position of power or experiencing the effects of presence/absence of power abuse. In addition, we use the same method to elicit normative beliefs of these three types of subjects *in other reference groups*, which allows us to test two theories that we have in regard to the nature of these normative perceptions.

The main hypotheses about the normative perceptions of situations arising in our version of Public Goods game are based on the Belief in a Just World theory (BJW) proposed by social psychologist Melvin J. Lerner and described in his book of the same title (Lerner, 1980). The main tenet of BJW is that people have a strong tendency to believe that "there is a pattern to events which conveys not only a sense of orderliness or predictability, but also the compelling experience of appropriateness expressed in the typically implicit judgment, "Yes, that is the way it should be." (Lerner, 1980, p. vii). In other words, this is a desire to maintain a coherent and orderly picture of reality in which good consequences are attributed to some individual's correct and appropriate actions, while bad ones to someone's inappropriate or ignorant behavior. The workings of BJW are mostly notable when random tragic events are "rationalized." For example, a widespread opinion after hurricane Katrina was that people, who suffered from the flood, were responsible since they chose to live on the land below sea level ("rejection of victims" as Lerner (1980) puts it). This view restores the belief in a just world, since otherwise one would have to admit that innocent people can be hurt for no reason. More importantly, similar rationalizations go through the minds of those who suffer: Lerner (1980) provides evidence that rape victims blame themselves in order to sustain an "illusory degree of control"; hostages and kidnapping victims defend the actions of their abusers, which came to be known as 'Stockholm syndrome' in popular press.[2] Finally, those who actually cause damage and suffering might also rationalize their own behavior: anecdotal evidence suggests that (at least some) Russian policemen are of an opinion that they "deserve" to extort bribes from ordinary citizens, since they themselves suffered from corrupt officers in the police academy.

Applied to our experiment, BJW predicts that punishers, who abuse power, should rationalize their free-riding and punishing behavior by believing that their actions are more socially appropriate than what punishers, who contribute as much or more than other players, believe.[3] Similarly, the same contributions and punishment choices of a punisher should be thought of as *more* appropriate by those who were exposed to the actions of an abusive punisher than by players who did not experience abuse.

Despite much evidence in support of BJW (Friesen *et al.*, 2018; Konow *et al.*, 2018), it might

[2] Namnyak *et al.* (2008) report that there is little published scientific research on 'Stockholm syndrome', though, they do not deny the existence of similarities among the cases depicted in the newspapers.

[3] In the context of our study we say that punisher abuses power when he enforces a norm for others to contribute some number of tokens to the public good, but himself contributes less. See Hoeft and Mill (2017) for discussion.

be the case that it is not that prominent in our experiment. After all, most examples of BJW involve direct and often rather serious physical or psychological harm, while modern economic power abuse is mostly indirect and is measured in monetary terms. Therefore, we propose an alternative hypothesis, which we coin the Belief in an Unjust World (BUW). It states that all punishers hold the same beliefs about the appropriateness of free-riding and punishing, and those, who abuse their power, understand that what they do is inappropriate, but simply do not care. Victims of the abuse also realize what is done to them and, having experienced it, believe that punishers' free-riding and punishment are *less* appropriate than players who did not experience abuse. Thus, for victims the predictions of BUW and BJW go in opposite directions.

We find strong support of BJW. In particular, those punishers who abuse power and those who do not indeed differ in their perception of the social norms related to free-riding: abusive punishers believe that it is more appropriate to contribute less than the amounts contributed by the powerless subjects. Players who experienced abuse also believe that it is more appropriate for punishers to free-ride *and to punish others* than players whose punisher contributed more than them. In fact, the norms elicited from abusive punishers and the victims of their abuse are statistically indistinguishable, as are the norms in groups where no abuse took place. Thus, the norms of the powerless converge to those of the powerful. We also find that punishers, regardless of their behavior, are of an opinion that outsiders, who did not experience the Public Goods game, share their normative beliefs. This shows that *simply being in power* already changes the ways people think about the appropriateness of their actions. However, we also find that both punishers and the powerless players think that normative beliefs in the opposite group are different in a way consistent with BUW. Given that the norms of abusive punishers and their victims are the same in their own reference groups, as are the norms of non-abusive punishers and the subjects who played with them, this result demonstrates that punishers and other players *do not notice* that their own beliefs were "corrected" by BJW. Thus, the beliefs about the norms of subjects in other roles, which are seemingly in line with BUW, are *wrong*.

These findings draw a rather grim picture in which the powerful abuse their position, believing that they do nothing wrong, the powerless suffer from the abuse, but consider their situation normatively appropriate. We show that BJW does have a strong effect on the normative perceptions related to indirect power abuse. If our results can be extrapolated to real economic environments, they can explain a relative stability of corrupt institutions, since no party involved is feeling that anyone is doing anything wrong. Indeed, in a recent World Bank report (World Bank Group, 2017) it is claimed that top-down attempts at fighting corruption fail due to social norms that support it on all levels of social hierarchy.

2 Experimental Design

To study the abusive behavior and the normative perception of power we conducted a two-part experiment. The first part is very similar to the design used in Hoeft and Mill (2017). In particular, a standard Public Goods game (PGG) is implemented for 15 rounds with one subject assigned to the additional role of punisher throughout the game. The second part of the experiment utilizes the design of Krupka and Weber (2013) to elicit subjects' normative perceptions of the game. More specifically, subjects in power, subjects not in power, and unrelated outsiders are asked to normatively evaluate several situations that could take place in PGG.

2.1 Public Goods Game

All participants were randomly assigned a fixed role, either punisher or non-punisher, and appointed to a group of four, in which they remained for the 15 rounds of the PGG (partner matching). Each round of PGG consists of three stages.

Stage 1. Contribution to the Public Good. The first stage is a standard PGG. Each of the four participants is endowed with 20 tokens and is asked to allocate this endowment between private and public accounts (1 token = 20 Euro cents). Tokens allocated to the private account are subject's to keep. Tokens allocated to the public account (c_i) have a marginal per-capita return (MPCR) of 0.5, so that each group member receives 0.5 times the total contribution. The payoff π_i of participant i is defined as

$$\pi_i = 20 - c_i + 0.5 \cdot \sum_{j \in \{1..4\}} c_j. \tag{1}$$

Stage 2. Punishment. In the second stage, the punishment decisions are made. While the three non-punishing group members (participants A, B, and C) are just shown a blank screen asking them to wait for the decision of the punisher, the punisher (participant D) is shown the contributions and current payoffs of all group members. The punisher is then asked to indicate how many points he would like to deduct from the payoff of subject i (σ_i, $i \neq D$).[4] To rule out reputation effects from previous rounds the information about non-punishing participants is presented to the punisher in random order in each round (Fehr and Gächter, 2000). The overall maximal possible deduction in every round is restricted to 30 tokens, which is enough to deter every participant from free-riding.[5] The punishment is costless for D and unused punishment tokens are

[4] To avoid framing and demand effects we referred to the act as "reducing the payoff" and not as "punishment."

[5] Note that the individual benefit of free-riding, compared to full contribution, is 10 tokens. If the punisher were confronted with three free-riders and utilized all 30 punishment tokens, he could make every free-rider indifferent between free-riding and fully contributing by subtracting 10 tokens from each of them. As soon as one subject contributes more than zero, the punisher can already make contributing a preferential option. Hence, 30 tokens are

forfeited.[6,7] Thus, the punisher could reduce the payoff of the non-punishers by 30 tokens at most, but his payoff would not be directly influenced by punishing (as punishment is costless) or not punishing (as unused tokens are forfeited). This is to ensure that the contributions of the punisher can be directly compared to the contributions of others.

The payoff π_i of a non-punisher $i \neq D$ is given by

$$\pi_i = 20 - c_i + 0.5 \cdot \sum_{j \in \{1..4\}} c_j - \sigma_i. \tag{2}$$

The payoff of the punisher is described by equation (1).

Stage 3. Feedback. The third stage provides feedback to the participants. More specifically, they are informed about their own contribution to the private and group accounts, their own punishment (reduction), and their resulting payoff. Further, they are also informed about the contributions of all other group members labeled as players A, B, C, and D throughout all rounds. Importantly, subjects are able to track the contribution behavior of the punisher. Non-punishers are not informed about the punishments meted out to others.

2.2 Norm Elicitation Task

To elicit normative perception we utilize the norm elicitation task by Krupka and Weber (2013). More specifically, subjects have to indicate how socially appropriate they find a certain action (five actions are assessed) in a certain situation (three situations are assessed). In order to be paid, participants are asked to indicate the *modal* appropriateness estimation of a specific group of other participants. If their assessment of the social appropriateness of a specific action in a specific situation in a specific group was identical to the assessment of the majority of other participants in this group, they were paid €8, otherwise they were paid €0. The three situations, with the corresponding five actions to be normatively assessed are as follows:

Question 20 Suppose the others (A, B, C) contributed 20 tokens each into the group account in the previous round. How socially appropriate are the following decisions by D?
D contributes 0, 5, 10, 15, 20 tokens to the group account.

Question 10 Suppose the others (A, B, C) contributed 10 tokens each into the group account in the previous round. How socially appropriate are the following decisions by D?
D contributes 0, 5, 10, 15, 20 tokens to the group account.

sufficient to ensure punishment to be a deterrent.

[6]Making punishment costly would change the budget constraint of the punisher, thus, making his contribution decisions incomparable to the contribution decision of the non-punishers.

[7]In the alternative case of not forfeiting punishment tokens, the punisher could contribute more in stage one, anticipating extra gains in the second stage, which again would make the contribution decisions of punishers and non-punishers incomparable.

Punishment Question Suppose the others (A, B, C) contributed 10 tokens each into the group account in the previous round. How socially appropriate is it for D to reduce the payoff of A, B, or C if he contributed the following amounts?
D contributes 0, 5, 10, 15, 20 tokens to the group account and reduces the payoff of A, B, or C.

In each of the three situations, subjects had to rate social appropriateness of each action (contribution by D of 0, 5, 10, 15, 20). For each action the appropriateness had to be chosen on a seven-point Lickert scale: very socially inappropriate, socially inappropriate, somewhat socially inappropriate, neither appropriate nor inappropriate, somewhat socially appropriate, socially appropriate, very socially appropriate.[8]

In the first task, to assess social appropriateness of these situations, punishers had to indicate what they think the majority of other punishers in the current session consider as socially appropriate (punishers' own reference group). Similarly, players A, B, and C had to indicate what they think the majority of other such players in the current session consider socially appropriate (ABCs' own reference group). Next we asked punishers what they think the majority of the non-punishers in the current session consider as socially appropriate. Similarly, we asked non-punishers what they think the majority of punishers in the current session consider as socially appropriate. We also asked both punishers and non-punishers what they thought the majority of a third group of people consider socially appropriate. This group were independent outsiders who did not participate in Part 1 of the experiment (PGG), but were given the same instructions as punishers and non-punishers. These subjects had to just indicate what they thought the majority of punishers, non-punishers, and other independent outsiders in their session consider socially appropriate.

Thus, there are three randomly assigned groups of people: punishers, non-punishers, and independent outsiders, who did not take part in PGG. All subjects in these groups had to first evaluate social appropriateness ratings of subjects in the same role. Then, subjects in each group evaluated social appropriateness in the other two groups.

2.3 Payment

At the end of the experiment, subjects were paid for three tasks: PGG, the appropriateness evaluation in their own reference group, and the guess of the appropriateness evaluation in other two reference groups.

1. Subjects in the role of punishers and non-punishers were paid for one randomly chosen round of PGG.

[8] We chose seven instead of five statements to reduce a possible demand effect, i.e., choosing different appropriateness level for each of the five actions. See Tables 4, 5, and 6 in Appendix A for further details.

2. One random action from one random situation of Part 2 was drawn to determine the payment. In case a subject evaluated the payoff-relevant action in the payoff-relevant situation as the majority of other subjects *in the same role* she obtained €8 and zero otherwise.

3. To determine the payoff for the guess of the appropriateness evaluation in other reference groups, one random situation and one random action was drawn in one random reference group. If a subject evaluated the payoff-relevant action in the payoff-relevant situation as the majority of others *in the randomly determined payoff-relevant group* she obtained €8 and zero otherwise.

Overall, the average payoff for punishers and non-punishers was €16.50 (including a show-up fee of €5). The average payoff for independent outsiders (who did not take part in PGG) was €9.30 (including the show-up fee).

2.4 Subjects

289 participants (60% female) were recruited with the online registration software Hroot (Bock *et al.*, 2014). The experiment was conducted at the Bonn DecisionLab and consisted of 9 sessions. The first session was run with 17 subjects who participated only in the second part and only in the appropriateness evaluation (not the guess of the appropriateness evaluation) to make further payments possible.[9] 7 sessions were conducted with participants in the roles of punishers and non-punishers (4 sessions with 32 subjects and 3 sessions with 28 subjects) and further 2 sessions with 30 participants each were conducted in the role of independent outsiders.

The participants' age ranged from 17 to 73 years (median = 22). Most were bachelor students (semester median = 3). The average earnings were €14.50 (including a €5 show-up fee). The experiment lasted 1.5 hours (including seating, instructions, payoff, etc.). All measurements were computerized with the experimental software z-Tree (Fischbacher, 2007).

Participants were randomly assigned to computer cubicles. They received written instructions separately and were given an opportunity to ask questions for each task in the experiment.[10] After taking part in PGG subjects were given on-screen instructions for the norm elicitation task and made their decisions in this task. After that, they filled in socio-demographic information and then were presented with their payoff information and received their payoff privately.

[9] To determine the payoff of punishers and non-punishers if their guess of the appropriateness evaluations of independent outsiders was deemed payoff-relevant we needed the actual appropriateness evaluation of this group.

[10] The instructions, as well as an English version of the handout, and the screenshots of the experiment, can be found in Appendix E.

3 Hypotheses and Predictions

In order to structure our results, we start with explicitly formulating the hypotheses about the norms elicited from subjects in the experiment within and between reference groups. Let us call subjects who played in role *D* in PGG *punishers*, subjects who played in roles *A*, *B*, and *C* *victims*, and subjects, who did not participate in PGG, *outsiders*. Since the focus of this study is on understanding the motives behind abusive behavior of punishers and its consequences, we will call punishers, who force victims in their groups to contribute more than themselves, *bad punishers*, and punishers, who contribute at par with or more than victims, *good punishers*. Respectively, *bad victims* are subjects in a group with a bad punisher, and *good victims* are those in a group with a good punisher. We will refer to these groups as *good* and *bad groups*.

In PGG punishers are free to choose any level of contribution and punishment in the sense that they are not influenced by punishment from other subjects. Victims, on the other hand, can be forced to contribute certain amount under threat of punishment. Therefore, it is reasonable to assume that punishers, if they care about following norms, will base their choices in PGG on what they perceive as socially appropriate. A particular experience during the game should not influence punishers' norms, since they are never coerced into choosing any specific action. Victims, however, can be forced to do punisher's bidding, which can be inconsistent with what they would have done if they could choose freely. Thus, their experience can have an effect on the perception of norms.

Following this argument, we will divide the hypotheses that we formulate into three classes. For punishers we will consider *behavioral hypotheses*, which describe punishers' norms and their choices in PGG. For victims we will formulate *experiential hypotheses*, which relate changes in victims' norms and their participation in good or bad groups. Finally, for punishers, victims, and outsiders, *belief hypotheses* will connect norms between reference groups.

We start with the behavioral hypotheses for the punishers. There are two possible ways they can perceive their own actions:

Hypothesis PJ *The Belief in a Just World (BJW) influences norm perception among punishers. Good punishers think that free-riding and punishing others is inappropriate, which rationalizes their behavior. Bad punishers, similarly, think that it is appropriate to free-ride and punish others, which also rationalizes their behavior.*

Hypothesis PU *According to the Belief in an Unjust World (BUW) there is a common perception of norms among punishers. Bad punishers, who violate these norms, understand what they are doing but do not care. Good punishers follow the norms.*

The consequences of these hypotheses for elicited norms are as follows. Under hypothesis PU, there is no reason to expect that bad punishers express norms differently from good punish-

ers in their own reference group. Conversely, under hypothesis PJ, we should observe that bad punishers consider free-riding and punishing *more* socially appropriate than good punishers do.

Next we consider experiential hypotheses for victims. There are two opposite ways in which experience can potentially shape victims' norms. According to BJW victims should adjust their norms to their experiences by starting to believe that the behavior of the punisher is justified. Thus, bad victims should think that punishment and free-riding are *more* appropriate than good victims do. According to BUW, bad victims should consider the acts of bad punishers as *less* appropriate than good victims, since they do experience abuse while good victims do not.

Hypothesis VJ *Victims change beliefs about norms with their experience in PGG in accordance with BJW. Bad victims think that free-riding and punishing are more appropriate than good victims do.*

Hypothesis VU *Victims change beliefs about norms with their experience in PGG in accordance with BUW. Bad victims think that free-riding and punishing are less appropriate than good victims do.*

Finally, when subjects guess what norms are expressed in other reference groups, it becomes important whether they *recognize* that BJW or BUW might have had an effect on their own beliefs and the beliefs of subjects who have other roles. In particular, if subjects understand the influence of BJW they should not think that the norms are different in other reference groups. However, if they understand BUW, then they *might* think that subjects in other reference groups have different beliefs about norms. Whether they do think this or not depends on subjects' experience. We postulate that subjects should only think that norms are different in other reference groups if 1) something is done to them that they find inappropriate and 2) they imagine that others might consider some of their actions inappropriate.

Hypothesis BJ *If subjects recognize the effects of BJW, everyone who participated in PGG and had similar experience should think that the norm is the same no matter the role, since beliefs about the norms adjust to experience.*

Hypothesis BU *If subjects recognize the effects of BUW, bad victims should think that punishers consider free-riding and punishing more appropriate than other victims. Punishers (good or bad) should think that victims consider free-riding and punishment less appropriate than other punishers. Outsiders should foresee both effects.*

4 Results

4.1 Good and Bad Groups

In this section we present the summary of results for PGG and explain the method of analyzing the data for elicited norms. In order to see whether punishers' norms shape their behavior we classify them according to their average contribution to the public good. Notice that for punishers the choice of how much to contribute is not constrained by punishment coming from other subjects. Moreover, this choice does not have to depend on the contributions of others, since a punisher can, in principle, force them to contribute any amount she wishes by applying punishment. Therefore, if punishers adhere to the norms in their choices, the average contribution should reflect this connection.

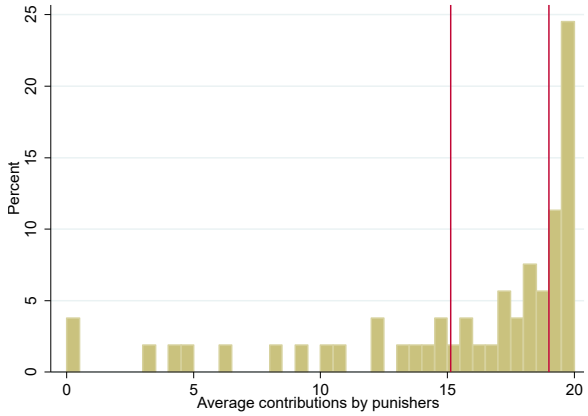

Figure 1: Histogram of average contributions by punishers divided into terciles.

Figure 1 shows the histogram of the average contributions by punishers (53 of them). The two red lines represent the division of the distribution into terciles. In most of the analyses that follow we will always compare choices in PGG groups that belong to either the bottom tercile (average contribution less or equal to 15.13) or the top tercile (average contribution greater or equal to 19).[11] In line with the previous section, we will refer to the groups of subjects from bottom and top terciles as *bad* and *good* groups with the corresponding adjectives for punishers and victims.

First, we look at the dynamics of contributions in good and bad groups. Figure 2 (left panel) shows the average contributions of good and bad punishers and victims.

There is a large difference in constributions of good and bad punishers. The former act very cooperatively and contribute on average more than victims in their groups. In addition, they apply punishment to increase the contributions of victims to their level, which is evident from the fact that good victims' contributions increase with time. Bad punishers contribute little themselves, but try to push the contributions of victims above their own contribution level. The

[11] We chose to use top and bottom terciles for expositional reasons. All main results go through with panusher's average contribution taken as a continuous variable.

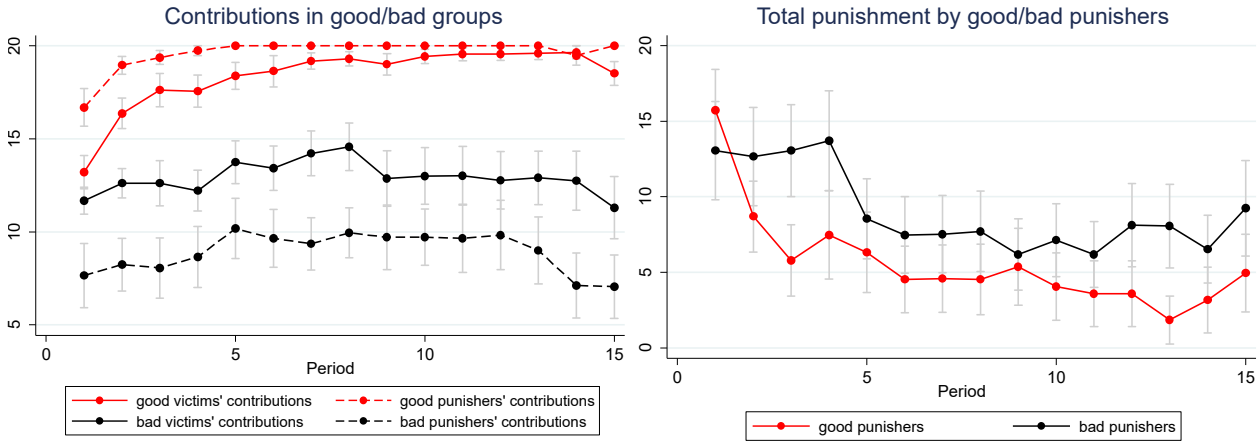

Figure 2: **Left panel.** Average contributions by punishers and victims in the top and bottom terciles of the distribution of punishers' average contributions. **Right panel.** Average total punishment in the same terciles.

right panel of Figure 2 shows that more punishment is used by bad than by good punishers, though, taken period-by-period and together, the amounts subtracted are not significantly different. Overall, we can conclude that the victims in the bad groups continuously feel that the cooperative norm imposed on them by their punishers keeps being violated by punishers themselves, and there is nothing victims can do about it. Conversely, in the good groups, punishers, if anything, serve as role cooperative models. This suggests that the victims in bad and good groups have rather different experiences and that it can have consequences for their perception of norms.

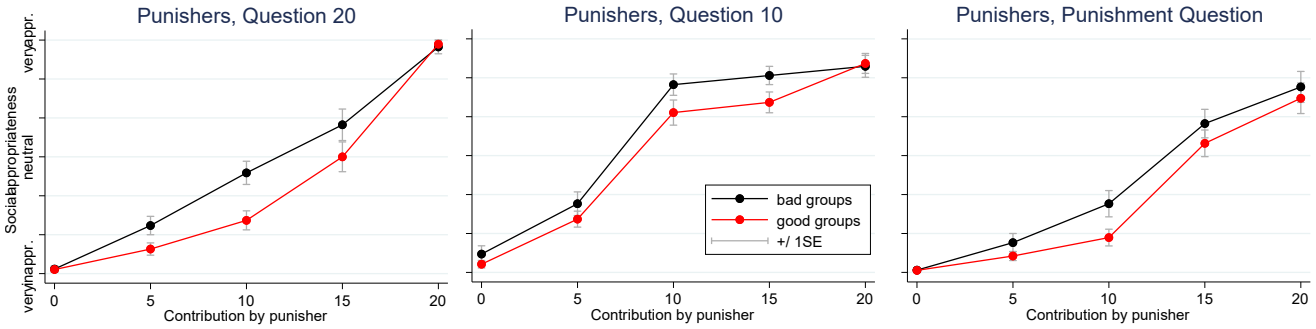

Figure 3: Norms expressed by good and bad punishers in their own reference group.

Next we look at the norms elicited with the norm elicitation tasks (Krupka and Weber, 2013). Figure 3 shows the norms expressed by good and bad punishers with the reference group being other punishers (own reference group). The leftmost graph shows the answers to Question 20. We see that everyone finds full contribution to be very socially appropriate and zero contribution to be very inappropriate. The difference between the two groups is noticeable for the intermediate answers: good punishers find it less appropriate than bad punishers to contribute intermediate amounts. From the perspective of the norm-dependent utility maximization (Kessler and Leider, 2012; Kimbrough and Vostroknutov, 2016; Thomsson and Vostroknu-

tov, 2017), this means that good punishers should make higher contributions than bad punishers since the derivative of the good punishers' norm is higher in the vicinity of full contribution, assuming that the norms are equal in the endpoints (hypothetical punishers' contributions of 0 and 20). The answers to Question 10 (middle graph) show that both bad and good punishers are conditional cooperators: they consider contributing anything above 10 tokens as approximately equally appropriate, but contributing less than 10 as inappropriate. If norm-dependent utility is maximized, such norm should lead to contributions of no more than 10, since contributing more decreases consumption payoff, but does not increase utility from norm compliance. Similarly to Question 20, good punishers consider it less appropriate to contribute any intermediate amount between 0 and 20, while at these endpoints the norms are roughly the same. By the above argument, this again implies that good punishers should contribute more than bad ones. Finally, the rightmost graph shows the answers to the Punishment Question. We see that all punishers agree that punishing after punisher contributed 0 herself is very inappropriate, but good punishers find punishing in general less appropriate than bad punishers. Since punishment does not influence the consumption utility of the punishers, this should lead to less punishment by good punishers than by bad punishers given the same contributions of others.

4.2 Norms and Behavior in Public Goods Game

In this section we demonstrate that the connection between the expressed norms and behavior is indeed in accordance with the norm-dependent utility maximization as we conjectured above. Notice that the norms, expressed by the participants in our experiment, are functions. Therefore, in order to compare them to contributions and punishment levels in PGG we need to transform them into single numbers. We consider the *average norms* over five levels of hypothetical punisher's contributions. Appendix B provides argumentation why this is a legitimate way to measure normative perceptions.

We start with punishers' attitudes towards free-riding, which are elicited by means of Question 20 in their own reference group. We expect that punishers' average contributions should be correlated with how socially appropriate they find different levels of contribution after victims contributed all 20 tokens in the previous period. The Spearman's rank correlation between the average contributions and the average norms is $\rho = -0.32$ ($p = 0.020$), which means that the lower the average norm, the higher is the average contribution, exactly as we predicted in the previous section. The linear regression in Table 1 (left column) shows that the average norm predicts average contribution (the descriptions of all variables used in the regressions can be found in Appendix C). The smallest average norm among punishers is 2.2 and the highest is 5. Thus, the regression predicts contributions in the interval $[9.7, 18]$, which means that the norms have a large influence on contributions.

For the norms expressed by punishers in the answers to Punishment Question, we find that

	Punishers'		Victims'		
Dependent variable:	average contribution	total punishment	average contribution		
Groups:	all	all	all	bad	good
pp-q20	–2.693** (1.162)				
pp-qpun		2.328** (1.154)			
vv-q20			–0.938*** (0.327)	–0.441 (0.719)	–1.346** (0.659)
constant	24.591*** (3.781)	–0.821 (3.552)	19.580*** (1.144)	14.485*** (3.055)	22.637*** (1.778)
N punishers/victims	53	53	159	51	57
N groups			53	17	19

Table 1: OLS regressions of punishers' average contributions and total punishment on the punishers' average norms in own reference group (left two columns). Random effects regressions of victims' average contributions on their average norms in own reference group (three right columns). Errors are robust (in case of random effects, also clustered by group). Standard errors in parentheses. * – $p < 0.1$; ** – $p < 0.05$; *** – $p < 0.01$.

they are positively correlated with the amount of total punishment (Spearman's $\rho = 0.29$, $p = 0.037$), which is again as we expected: the more appropriate the punishment, the more of it is being used. The regression in Table 1 (second column) also supports this finding. For the range of average punishment norms $[1, 4.4]$ the regression predicts total punishment in the interval $[1.5, 9.4]$, which, again, is non-negligible.

These results for punishers demonstrate that the expressed norms do correlate with behavior as we predicted.[12] However, given our interest in testing the hypotheses related to BJW, which are inherently about changing beliefs, it is important to mention at this point that we do not think that punishers necessarily change their perception of the norms *during* PGG or rationalize their behavior *after* they choose in PGG, as Hypothesis PJ may seem to suggest. Instead, we hypothesize that each punisher, given her tendencies to behave in situations similar to punisher's role in PGG, already comes to the lab "equipped" with the beliefs about norms that rationalize her choices, as BJW would predict. Therefore, in our opinion, the correlation between punishers' contributions/total punishment and norms reflects the already existing connection and not something acquired during PGG. In this respect, the results above provide first evidence of the presence of BJW in punishers.

For victims, though, the situation is different since during PGG they are influenced by the punishment inflicted upon them. Thus, we conjecture that victims' contribution levels should not correlate with their norms. Nevertheless, we do find that victims' average norms in Question 20 (in own reference group) are correlated with their average contributions (Spearman's $\rho =$

[12]The correlation for Question 10 is as for Question 20, but is only weakly significant at 10% level.

-0.37, $p < 0.001$, middle column in Table 1). However, if our hypothesis that victims' norms are influenced by their experience in PGG is correct, this correlation can be an artifact of changing perceptions of norms. Indeed, according to Hypothesis VJ, bad victims, who have experienced bad punishers, should have higher average norms in Question 20 than good victims, who have experienced good punishers, and, as we already know from Figure 2, good victims contribute much more than bad victims, which can explain the connection. We calculate the correlation coefficients for bad groups and good groups separately and find that in the former it is $\rho = -0.23$ ($p = 0.090$) and in the latter $\rho = -0.27$ ($p = 0.038$). Similar results are obtained by the regression analysis (two right columns in Table 1). So, indeed, bad victims' contributions are not correlated with their expressed norms, but good victims' contributions still are. This means that victims in good groups are not constrained by punishment as much as victims in bad groups.

Result 1. *The norms expressed by punishers do reflect how much they contribute to the public good and how much they punish the victims, which constitutes first evidence of Hypothesis PJ. The contributions of victims are correlated with the norms only in good groups. In bad groups the correlation is weak or non-existent, which is probably due to the influence of punishment.*

4.3 Punishers' Norms

Result 1 provides some support for Hypothesis PJ. However, the existence of a correlation does not guarantee that there is a significant difference in norms between good and bad groups.

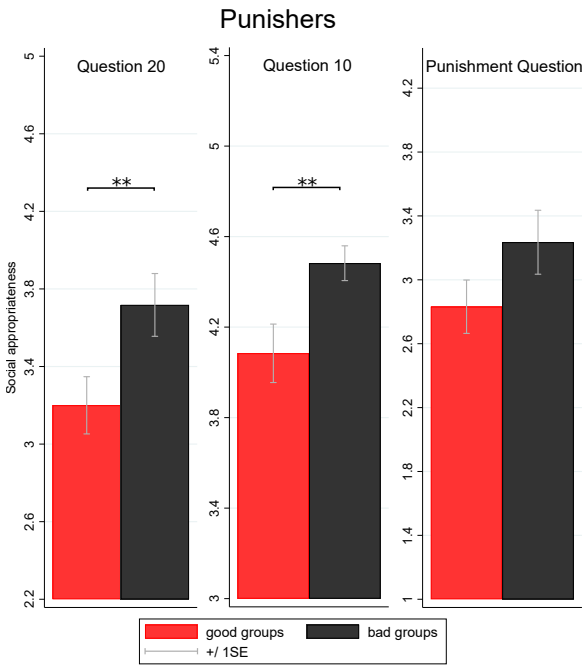

Figure 4: Average punishers' norms in own reference group. For each question, y-axis ranges from minimum to maximum value of the corresponding average norm. Significance levels of the permutation tests of means are reported. * – $p < 0.1$; ** – $p < 0.05$; *** – $p < 0.01$.

Figure 4 shows punishers' average norms for the three questions in their own reference group (Figure 3 shows same norms as functions). We see that there is a significant difference in the average norms between good and bad groups for Question 20 (permutation test, $p = 0.025$): bad punishers consider it more appropriate than good punishers to free-ride after others contributed full amount.[13] Similar difference can be observed for Question 10 (permutation test, $p = 0.016$): bad punishers consider it more appropriate than good punishers to contribute small amounts after victims contributed 10 tokens in the previous period. For the Punishment Question the difference is not significant (permutation test, $p = 0.132$). Table 8 in Appendix D reports similar results for all punishers (not only good and bad ones). These findings support Hypothesis PJ and BJW: bad punishers justify their behavior to themselves by believing that contributing little is not that bad from the moral perspective.[14]

Result 2. *Punishers' norms are in line with Hypothesis PJ. Bad punishers justify their free-riding by believing that their behavior is socially appropriate. Good punishers contribute a lot because they think that doing otherwise is inappropriate.*

4.4 Victims' Norms

All our results concerning punishers' norms stemmed, to some extent, from the fact that we divided the groups into bad and good according to punishers' average contributions. This, however, is not true for victims, who were assigned randomly to good and bad groups. Therefore, any differences in norms that we detect between good and bad victims must be due to the experience that they had during PGG. This gives us an opportunity to see how the oppressive and corrupt behavior of bad punishers and cooperative behavior of good punishers changes victims' perception of the appropriateness of punishers' actions.

Figure 5 shows the average victims' norms in their own reference group (Figure 9 in Appendix D shows same norms as functions). Answers to Question 20 tell us what victims believe is the common attitude among victims towards *punishers'* free-riding. We see that bad victims consider it significantly *more* appropriate than good victims (permutation test, $p = 0.002$). Similarly to punishers, this result is in support of Hypothesis VJ and BJW: bad victims justify the low contributions of punishers by believing that this is socially appropriate. The left column

[13] Here and below all tests are two-tailed. We chose permutation test over rank-sum test since the latter is not a test of difference in means, but difference in distributions. Therefore, it can be significant even when the means are not statistically different. The permutation test that we use is a test of difference in means.

[14] Kimbrough and Vostroknutov (2018) make similar observations about the sharing behavior of *rule-followers* and *rule-breakers* in Dictator game: the latter give less than the former and consider it more appropriate to do so. We did not include the rule following task in our experiment due to time constraint, however, the multitude of previous results that connect the preference for rule-following and cooperative behavior suggest that similar connection may exist here as well. Therefore, it is probable that bad punishers behave selfishly because they actually do not care about following norms, but, in line with BJW, still hold a belief that what they do is not inappropriate (see more discussion in Section 5).

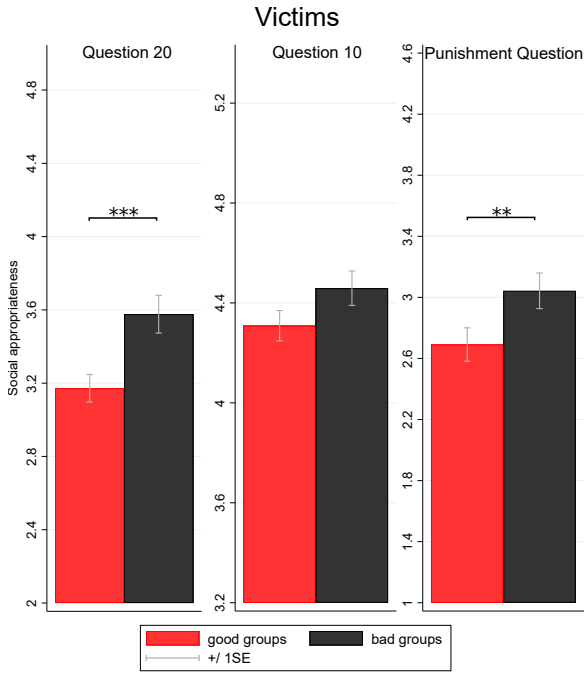

Figure 5: Average victims' norms in own reference group. For each graph, y-axis ranges from minimum to maximum value. Significance levels of the permutation test of means are reported. * – $p < 0.1$; ** – $p < 0.05$; *** – $p < 0.01$.

in Table 9, Appendix D, demonstrates the same point with a regression and a rank correlation that use all data instead of only good and bad groups. Notice that, unlike Section 4.2, here we show how victims' norms depend on *punishers'* average contributions, and not on victims' own contributions.

Figure 5 shows that bad victims also consider it significantly more appropriate than good victims when punishers subtract money from others (same results for all data: right column in Table 9, Appendix D). Importantly, unlike punishers, *victims are not those who punish, but those who receive the punishment*. Therefore, bad victims, instead of seeing the hypocritical punishment, which comes from a person who contributes less than them, as "unfair" and, thus, inappropriate (Hypothesis VU), start to believe that it is actually justified (Hypothesis VJ). These two results demonstrate an astounding effect that negative experiences can have on the perception of appropriateness.

Result 3. *Victims' norms are in line with Hypothesis VJ. Bad victims see low contributions of the punishers and the punishment that they receive as more appropriate than good victims.*

4.5 Comparison of Victims' and Punishers' Norms

We have seen that victims' norms are modulated by the experience in PGG, and that the bad victims' norms are higher than good victims' ones, exactly same relationship that we found between punishers' norms in good and bad groups. The next logical step is to test if the norms

of victims and punishers are similar in good and bad groups.

Group:		Good	Bad
Question 20	Victims	3.172	3.576
	Punishers	3.200	3.717
Question 10	Victims	4.309]*	4.458
	Punishers	4.084	4.482
Punishment Question	Victims	2.691	3.043
	Punishers	2.831	3.235

Table 2: Average norms of victims and punishers in own reference groups. No significant differences, all permutation tests $p > 0.423$, except good groups comparison for Question 10 ($p = 0.094$).

Table 2 shows victims' and punishers' average norms in their own respective reference groups. We see that the norms, when considered separately in good and bad groups, are not significantly different from each other (except for one comparison with $p = 0.094$). For example, for Question 20 good victims' average norm is 3.127 and good punishers' 3.200, which are almost identical. Similarly, bad victims' and punishers' norms are 3.576 and 3.717. This clearly demonstrates that victims' norms in good and bad groups have converged to the norms of good and bad punishers, which is perfectly consistent with BJW.

Result 4. *Victims' norms in good and bad groups converge to the norms of good and bad punishers, which is consistent with BJW.*

4.6 Norms between Reference Groups

In order to test the belief hypotheses stated in Section 3, we analyze norms between reference groups. This, however, cannot be done by simply comparing average norms in own and other reference groups. The reason is that subjects, when they decide which norms are prevalent in other groups, might be biased by the norm that they think exists in their own reference group.[15] It can happen that a norm that a subject thinks is in place in his own group weighs in her judgement about the norms in other groups.

Table 3 shows Spearman's rank correlations of norms between own reference group and two other reference groups. The correlations are, indeed, rather high for all questions and groups. Thus, it is true that subjects, when assessing what norms other reference groups may have, rely heavily on the individual perception of the norm in their own reference group. It does not necessarily mean that they do not understand that other subjects might have different ideas about

[15]For example, Eijkelenboom *et al.* (2018) find that in a social responsibility experiment, where subjects make risky choices for others, those with extreme risk preferences think that the average risk attitude in the population is much closer to their own risk preference than it actually is. Their own risk preferences bias their estimates of the population average.

Group:	Punishers		Victims		Outsiders	
Reference group:	Victims	Outsiders	Punishers	Outsiders	Victims	Punishers
Question 20	0.776	0.821	0.610	0.706	0.672	0.460
Question 10	0.602	0.647	0.400	0.465	0.440	0.341
Punishment Question	0.561	0.620	0.453	0.590	0.816	0.615
N subjects	53	53	159	159	59	59

Table 3: Spearman's rank correlations of average norms between own reference group and other two reference groups. All $p < 0.001$.

what is socially appropriate. However, this does imply that subjects with extreme opinions about the prevailing norm will under/over-estimate how distant they are from the average opinions about social appropriateness.

In order to estimate the "true" norm that subjects think is present in other groups, we propose a method of de-biasing the expressed norms. Suppose that subject i of type τ (punisher, victim, or outsider) expresses average norm x_i in her own reference group. Assume also that there is a true average norm g_τ that all subjects of type τ try to express when guessing the norm in some other group. However, subject i is biased in that instead of expressing g_τ she expresses some convex combination $y_i = \alpha_\tau x_i + (1 - \alpha_\tau) g_\tau$, which we observe. The problem now is to find estimates of g_τ and α_τ from known pairs (y_i, x_i). Let us regress y_i on x_i and obtain the parameters of the linear fit: $y_i = c + b x_i$, where b and c are the coefficients from a linear regression. Then, g_τ and α_τ are easily expressed in terms of b and c as $\alpha_\tau = b$ and $g_\tau = c/(1 - b)$. Thus, all we need to do is to run linear regressions of norms expressed by subjects in other groups on the norms from their own group and calculate α_τ and g_τ for each case.

Table 10 in Appendix D shows the regressions of average norms in victims'/punishers' reference groups on the average norm in own reference group for punishers/victims. Each regression estimates single parameter $b = \alpha_\tau$ (coefficient on the variable average norm) and different intercepts c. The coefficients on variable average norm, the estimates of α_τ for each question, are rather high. However, we are interested in the estimates of g_τ and how they compare to the average norms that punishers and victims express in their own groups.

Figure 6 shows the values of g_τ minus the average norm in own reference group for punishers and victims.[16] We see that both victims and punishers think that norms in the other group are different. Thus, we can reject Hypothesis BJ that subjects understand the influence of BJW. Notice that the direction of changes in appropriateness for both victims and punishers is exactly in line with Hypothesis BU. Punishers think that victims consider free-riding and punishing (Question 20 and Punishment Question) less appropriate than themselves, and victims think the opposite.

Next, we look at the choices of outsiders, who did not choose in PGG. Table 13 in Appendix D shows the regressions of outsiders' norms in punishers' and victims' groups on their own

[16] We do not consider bad and good groups separately since the estimates are roughly the same for both. See Figure 10 in Appendix D.

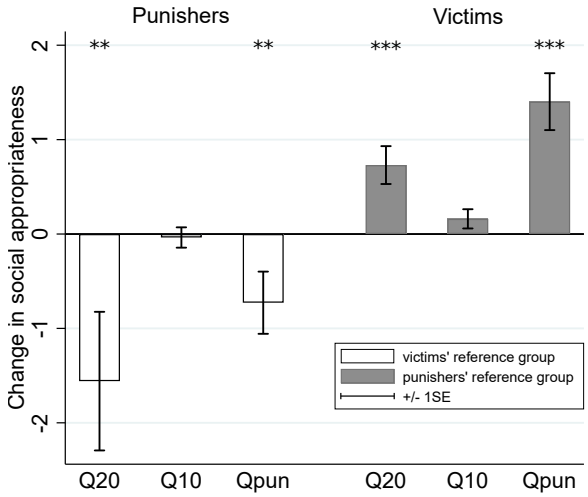

Figure 6: Estimates of g_τ minus the average norms in own reference group for punishers and victims. Above/below zero values mean that punishers/victims think that victims/punishers consider actions in a given question more/less socially appropriate than they themselves do in their own reference group. ***, **, * denote statistical significance at the 1, 5, and 10 percent level.

norms. As before the estimates of α_τ are significant and large, though for Questions 20 and 10 they are smaller than those that punishers and victims have.

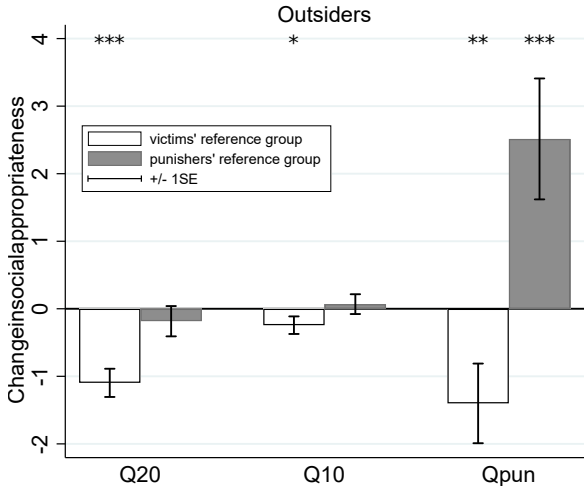

Figure 7: Estimates of g_τ minus the average norms in own reference group for outsiders. ***, **, * denote statistical significance at the 1, 5, and 10 percent level.

Figure 7 shows the differences between the estimates of g_τ's and respective average norms in outsiders' own reference group. In accordance with Hypothesis BU, we see that outsiders think that victims consider it less appropriate to free-ride and punish than themselves and the opposite for punishers. Overall, the changes in outsiders' beliefs are similar to those of punishers' and

victims' in Figure 6 (compare white and grey bars on the two figures).

Result 5. *We find strong support for Hypothesis BU and reject Hypothesis BJ.*

It may seem strange that we find support for BJW when we look at subjects' beliefs in own reference group, but reject BJW in favor of BUW when subjects' beliefs about other reference groups are concerned. This is an important finding, but we leave its discussion for Section 5 and continue with the last piece of evidence regarding punishers' and victims' beliefs about outsiders' norms. Notice that in this case both punishers and victims are asked the exact same question. Thus, the differences that we might observe should come from the assigned roles.

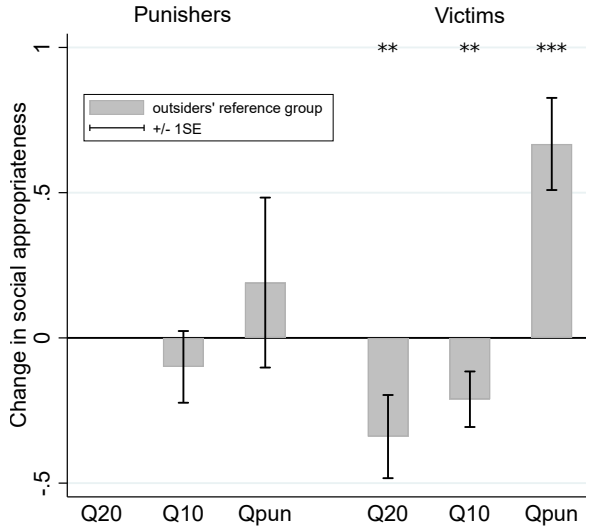

Figure 8: Estimates of g_τ in outsiders' reference group minus the average norms in own reference group for punishers and victims. ***, **, * denote statistical significance at the 1, 5, and 10 percent level.

We see from Figure 8 that punishers do not show any significant deviations from the norms in their own reference group when asked about the norms among outsiders.[17] For example, for Question 20 the coefficient on average norm is 0.977 and intercept is insignificant (leftmost column of Table 11 in Appendix D). Thus, most punishers are just repeating the norm that they expressed in their own reference group. As a result, the estimate of g_τ is very large, negative, and not significant (≈ -10, not shown on the graph). Similarly for Question 10 and Punishment Question we do not detect any significant difference between punishers' estimates in own reference group and in outsiders' group.

For victims, however, the picture is different. They think that outsiders consider free-riding less appropriate than themselves, and punishing more appropriate. Victims realize that outsiders have never played PGG, which means that they never experienced free-riding that is ubiquitous in PGG. Thus, victims think that outsiders might have an opinion that free-riding is worse than what victims think themselves. At the same time, victims experience punishment

[17] The same graph for bad and good groups separately is shown on Figure 11 in Appendix D.

and realize how bad it is as compared to outsiders who never experienced it. So, victims express this difference, which is consistent with BUW.

Result 6. *Punishers think that outsiders have the same norms as themselves. Victims' beliefs about outsiders are consistent with BUW.*

5 Discussion

Summary of the Results. The six results above provide a coherent picture of how the ability to abuse power influences punishers, victims, and their beliefs about the appropriateness of abusive behavior. The power over others has a significant influence on the social beliefs of punishers. Those who actually choose to abuse their power convince themselves that they are not violating any norms by doing so, while punishers who contribute more than others believe that abusing power is inappropriate (Result 2). These observations are in line with BJW, which can also explain what happens to those who are subject to the actions of a powerful person. Victims' beliefs about appropriateness of free-riding and punishment are changed by their experience in PGG (Result 3) and converge to those of their punishers, good or bad (Result 4). This convergence can be seen as a defensive mechanism that restores a meaningful world view when unfair circumstances cannot be changed (Lerner, 1980).[18]

However, when we analyze the beliefs about norms in other reference groups, we find that punishers hold an opinion that victims consider free-riding and punishment less appropriate, while victims think that punishers consider them more appropriate (Result 5). This, in principle, is consistent with BUW, which states that everyone understands what is done to them or what they do to others and how it is perceived. Nevertheless, we know that punishers' and victims' beliefs in their own reference groups are *identical* due to BJW (Result 4). Therefore, it seems that both victims and punishers are *not aware* that BJW belief adjustment takes place, which results in all of them having *wrong* beliefs about the other reference group. Thus, even though BUW formally manifests itself in the beliefs of subjects about other reference groups, this hypothesis is not consistent with these subjects' beliefs in their own reference groups and should, therefore, be rejected as own-beliefs-forming hypothesis in favor of BJW. This being said, we still think that BUW has its merits, since subjects use it to construct beliefs about normative perception of power abuse in other groups of people. Despite these beliefs being wrong, they can, nevertheless, reveal themselves through actions with tangible consequences.

Finally, we find a significant difference in how punishers and victims express their beliefs about the norms of the outsiders (Result 6). Punishers think that outsiders' norms are the same as their own, which suggests that *just being assigned to a position of power* convinces them that

[18]Importantly, Lerner (1980) also demonstrates that when victims of unfair treatment or outside observers *do have the means* to punish wrongdoing, their beliefs *do not* adjust in the direction of justifying such behavior.

what they do, abusing the power or not, is "right" in the eyes of outside observers. Such self-deception can lie at the core of the mechanism that sustains power abuse. At the same time, victims are sensitive to the fact that outsiders, who did not directly experience the actions of the powerful, might have different opinion about the appropriateness of punishers' choices. This further strengthens the conclusion that the powerful use any means to justify their behavior to themselves.

BJW and Rule-Following Propensity. From many studies (e.g., Kimbrough and Vostroknutov, 2016; Gürdal *et al.*, 2018) we know that the propensity to follow rules correlates with pro-social behavior. This means that rule-followers exhibit cooperative tendencies supported by the corresponding norms, while rule-breakers act selfishly. Theoretically, a selfish agent, who maximizes her own payoff in a role of a punisher in our PGG should contribute nothing and push others to contribute full amounts. This is very close to the behavior of bad punishers that we observe. Thus, there are two explanations for the bad punishers' behavior. First one is that bad punishers are rule-breakers, no matter their beliefs, and second is the one that we proposed, namely, that bad punishers think that free-riding is not inappropriate, no matter what their rule-following propensity is.

Our design does not allow us to cleanly separate which of the two factors, rule-following propensity or beliefs, drive bad punishers' behavior. However, the result about norms elicited in Dictator game presented in Figure 4 of Kimbrough and Vostroknutov (2018) suggests that there is a connection between being a rule-breaker and believing that behaving selfishly is appropriate. In particular, rule-breakers tend to think that selfishness is more appropriate than rule-followers do. If the same is true in our setting then bad punishers should be mostly rule-breakers, or selfish individuals, who think that free-riding is appropriate. Thus, the two explanations for abusive behavior might not be mutually exclusive, but actually constitute one explanation: inherently selfish individuals, who, nevertheless, are not exempt from the influence of BJW, rationalize their selfishness by believing that acting asocially is appropriate, while norm abiding individuals reinforce their prosocial behavior by believing that it is inappropriate to do otherwise. Additional experiments are needed to confirm or disconfirm this hypothesis.

Comparison to the Broken Windows Theory (BWT). The broken windows theory, which found certain experimental support (e.g., Funk and Kugler, 2003; Corman and Mocan, 2005; Engel *et al.*, 2014), states that when people see the results of others' not following norms (broken windows that stay unfixed) they also stop following norms *in other domains* thus hurting the community. This might sound similar to our results, however, there is a conceptual difference. BWT focuses on the idea that the appearance of run-down communities, which are not properly maintained, sends a signal that bad behavior stays unpunished, thus, giving a license to not follow norms. This does not mean that individuals, who break norms in these circumstances, start considering such behavior appropriate. Indeed, it might well be that, when they move to an appropriately

maintained neighborhood, they start to behave accordingly. Thus, BWT does not make any claims with regard to the *change in normative perception* that we emphasize in our results.

What we find is, in a sense, more serious than the effect of BWT. This can be illustrated with the example of bad victims who, after experiencing free-riding on the part of the punisher and her unfair punishment, start to believe that *the majority of other victims think* that such acts are normatively justifiable. Notice that these are the subjects who actually suffer from the abuse of power. Nevertheless, they start to share the viewpoint of bad punishers on such behavior. This suggests that corruption can breed more corruption even among those who never exercised but just experienced it. Undoubtedly, with our results we cannot support this statement or make any claims about how deep and lasting the effect of bad victims' negative experience is. However, we hope that our study can be the first step on the path to better understand these issues.

6 Conclusion

We study normative perceptions of power abuse in an experiment where only one player in a repeated Public Goods game (punisher) has a power to punish others conditional on their contributions. After the Public Goods game we measure the beliefs of all subjects about the appropriateness of punisher's actions by means of a norm elicitation task (Krupka and Weber, 2013). We hypothesize that the beliefs of the punishers and their victims are influenced by the Belief in a Just World (BJW, Lerner, 1980), a theory that states that people have a strong desire to maintain a coherent picture of the world in which good behavior is praised, bad behavior is punished or non-existent, and there is no place for wrongful acts that do not have retributive consequences. In line with BJW, we find that punishers, who abuse their power by contributing little and forcing others to contribute a lot, hold beliefs that this behavior is appropriate, while punishers, who contribute more than others, believe that abusing power is inappropriate. Other players, who experience the actions of the powerful, start to believe that these actions are justified no matter how abusive they are. More importantly, we find that neither punishers nor other players notice that their beliefs about the norms are getting influenced in this way. Our results unveil a mechanism that might be responsible for many failed attempts to fight corruption on an international level and point towards a reason why inefficient institutions endure.

References

ACEMOGLU, D., NAIDU, S., RESTREPO, P. and ROBINSON, J. A. (2015). Democracy, redistribution, and inequality. In *Handbook of income distribution*, vol. 2, Elsevier, pp. 1885–1966.

— and ROBINSON, J. A. (2008). Persistence of power, elites, and institutions. *American Economic Review*, **98** (1), 267–93.

BANERJEE, R. (2016). Corruption, norm violation and decay in social capital. *Journal of Public Economics*, **137**, 14–27.

BECKER, S. O., BOECKH, K., HAINZ, C. and WOESSMANN, L. (2015). The empire is dead, long live the empire! Long-run persistence of trust and corruption in the bureaucracy. *The Economic Journal*, **126** (590), 40–74.

BEETHAM, D. (2013). *The legitimation of power*. Macmillan International Higher Education.

BOCK, O., BAETGE, I. and NICKLISCH, A. (2014). Hroot: Hamburg Registration and Organization Online Tool. *European Economic Review*, **71** (C), 117–120.

CORMAN, H. and MOCAN, N. (2005). Carrots, sticks, and broken windows. *The Journal of Law and Economics*, **48** (1), 235–266.

CUBITT, R. P., DROUVELIS, M., GÄCHTER, S. and KABALIN, R. (2011). Moral judgments in social dilemmas: How bad is free riding? *Journal of Public Economics*, **95** (3?4), 253–264.

DAL BÓ, E., DAL BÓ, P. and SNYDER, J. (2009). Political dynasties. *The Review of Economic Studies*, **76** (1), 115–142.

DI TELLA, R., PEREZ-TRUGLIA, R., BABINO, A. and SIGMAN, M. (2015). Conveniently upset: avoiding altruism by distorting beliefs about others' altruism. *American Economic Review*, **105** (11), 3416–3442.

EIJKELENBOOM, G., ROHDE, I. and VOSTROKNUTOV, A. (2018). The impact of the level of responsibility on choices under risk: the role of blame. *Experimental Economics*, **forthcoming**.

ENGEL, C., BECKENKAMP, M., GLÖCKNER, A., IRLENBUSCH, B., HENNIG-SCHMIDT, H., KUBE, S., KURSCHILGEN, M., MORELL, A., NICKLISCH, A., NORMANN, H.-T. *et al.* (2014). First impressions are more important than early intervention: qualifying broken windows theory in the lab. *International Review of Law and Economics*, **37**, 126–136.

FEHR, E. and GÄCHTER, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, **90** (4), 980–994.

— and SCHURTENBERGER, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, **2** (7), 458–468.

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

FISMAN, R. and MIGUEL, E. (2007). Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets. *Journal of Political Economy*, **115** (6), 1020–1048.

FRIESEN, J. P., LAURIN, K., SHEPHERD, S., GAUCHER, D. and KAY, A. C. (2018). System justification: Experimental evidence, its contextual nature, and implications for social change. *British Journal of Social Psychology*.

FUNK, P. and KUGLER, P. (2003). Dynamic interactions between crimes. *Economics Letters*, **79** (3), 291–298.

GÄCHTER, S. and SCHULZ, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, **531** (7595), 496–499.

GLAESER, E. L., SACERDOTE, B. and SCHEINKMAN, J. A. (1996). Crime and social interactions. *The Quarterly Journal of Economics*, **111** (2), 507–548.

GREENE, J. D., CUSHMAN, F. A., STEWART, L. E., LOWENBERG, K., NYSTROM, L. E. and COHEN, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, **111** (3), 364–371.

GÜRDAL, M. Y., TORUL, O. and VOSTROKNUTOV, A. (2018). Norm compliance, enforcement, and the survival of redistributive institutions, mimeo, Boğaziçi University and University of Trento.

HERZ, H. and TAUBINSKY, D. (2017). What Makes a Price Fair? An Experimental Study of Transaction Experience and Endogenous Fairness Views. *Journal of the European Economic Association*, pp. 1–37.

HOEFT, L. and MILL, W. (2017). Abuse of power–an experimental investigation of the effects of power and transparency on centralized punishment, mimeo, University of Mannheim and MPI Bonn.

KESSLER, J. B. and LEIDER, S. (2012). Norms and contracting. *Management Science*, **58** (1), 62–77.

KIMBROUGH, E. and VOSTROKNUTOV, A. (2016). Norms make preferences social. *Journal of European Economic Association*, **14** (3), 608–638.

— and — (2018). A portable method of eliciting respect for social norms. *Economics Letters*, **168**, 147–150.

KONOW, J., JOHANSSON-STENMAN, O., MARTINSSON, P. and MEDHIN, H. (2018). The Just World Hypothesis: Theory and a natural field experiment, working Paper, Loyola Marymount University.

KRUPKA, E. L. and WEBER, R. A. (2013). Identifying social norms using coordination games: why does dictator game sharing vary? *Journal of European Economic Association*, **11** (3), 495–524.

LERNER, M. J. (1980). *The belief in a just world*. Springer Science+Business Media New York.

LOWES, S., NUNN, N., ROBINSON, J. A. and WEIGEL, J. L. (2017). The Evolution of Culture and Institutions: Evidence From the Kuba Kingdom. *Econometrica*, **85** (4), 1065–1091.

NAMNYAK, M., TUFTON, N., SZEKELY, R., TOAL, M., WORBOYS, S. and SAMPSON, E. L. (2008). 'Stockholm syndrome': psychiatric diagnosis or urban myth? *Acta Psychiatrica Scandinavica*, **117** (1), 4–11.

REUBEN, E. and RIEDL, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, **77** (1), 122–137.

ROSE-ACKERMAN, S. and PALIFKA, B. J. (2016). *Corruption and government: Causes, consequences, and reform*. Cambridge university press.

SCHROEDER, D. A. and GRAZIANO, W. G. (2015). *The Oxford Handbook of Prosocial Behavior*. Oxford ; NewYork: Oxford Library of Psychology.

TABELLINI, G. (2008). Institutions and culture. *Journal of the European Economic Association*, **6** (2-3), 255–294.

— (2010). Culture and Institutions: Economic Development in the Regions of Europe. *Journal of the European Economic Association*, **8** (4), 677–716.

THOMSSON, K. and VOSTROKNUTOV, A. (2017). Small-world conservatives and rigid liberals: Attitudes towards sharing in self-proclaimed left and right. *Journal of Economic Behavior and Organization*, **135**, 181–192.

WILSON, J. Q. and KELLING, G. L. (1982). Broken windows. *Atlantic monthly*, **249** (3), 29–38.

WORLD BANK GROUP (2017). *World Development Report 2017 : Governance and the Law*. Washington, DC: World Bank.

ZENOU, Y. (2003). The Spatial Aspects of Crime. *Journal of the European Economic Association*, **1** (2-3), 459–467.

Appendix (for online publication)

A Details of the Design

Suppose the others (A, B, C) contributed **20** tokens each into the group account in the previous decision.
How socially appropriate are the following decisions by D?

	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Neither appropriate nor appropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
D contributes 0 tokens to the Group account	✓						
D contributes 5 tokens to the Group account		✓					
D contributes 10 tokens to the Group account		✓					
D contributes 15 tokens to the Group account		✓					
D contributes 20 tokens to the Group account						✓	

Table 4: Example of norm elicitation, Question 20.

Suppose the others (A, B, C) contributed **10** tokens each into the group account in the previous decision.
How socially appropriate are the following decisions by D?

	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Neither appropriate nor appropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
D contributes 0 tokens to the Group account	✓						
D contributes 5 tokens to the Group account		✓					
D contributes 10 tokens to the Group account						✓	
D contributes 15 tokens to the Group account							✓
D contributes 20 tokens to the Group account							✓

Table 5: Example of norm elicitation, Question 10.

Suppose the others (A, B, C) contributed **10** tokens each into the group account in the previous decision. How socially appropriate is it for D **to reduce the payoff of A, B, or C** if he contributed the following amounts?

	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Neither appropriate nor appropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
D contributes 0 tokens to the Group account and reduces the payoff of A, B, or C.	✓						
D contributes 5 tokens to the Group account and reduces the payoff of A, B, or C.		✓					
D contributes 10 tokens to the Group account and reduces the payoff of A, B, or C.				✓			
D contributes 15 tokens to the Group account and reduces the payoff of A, B, or C.					✓		
D contributes 20 tokens to the Group account and reduces the payoff of A, B, or C.						✓	

Table 6: Example of norm elicitation, Punishment Question.

B Average Norms and Comparison of Endpoints

In our analysis we compare norms within and between subjects. In particular, for each subject, each question, and each reference group we compute the *average norm* with average taken over five levels of potential contributions of a punisher. Suppose we choose to compare the norms between two groups of subjects. For Question 20, *if the norms in these two groups are the same at the endpoints* (hypothetical punisher's contributions of 0 and 20), then the average norm becomes a measure of convexity of the norm function, or, in other words, the measure of steepness of the derivative in the vicinity of full contribution. For example, in the left panel of Figure 3, the average norm in the good group is smaller than the average norm in the bad group. With the endpoints assumption, this implies that lower average norm is equivalent to having steeper derivative close to full contribution, or, higher contributions according to the norm-dependent utility maximization. Similar argument holds for Question 10. For the Punishment Question the logic is slightly different: punishers do not incur costs when they choose how much to punish, so, in this case, lower average norm should automatically imply less punishment.

In order to compare norms in this way we need to show that for Questions 20 and 10 it is indeed the case that the norms at the endpoints are the same for all groups of subjects that we consider. This appendix provides the details of the statistical comparison of endpoints for the groups of subjects that we are interested in: good/bad punishers, good/bad victims and outsiders. With few exceptions, which do not undermine our arguments, we show that there are no reasons to suspect that the endpoints in our groups of interest are different. Therefore, it is legitimate to conduct all analyses using average norms.

We test the hypotheses that for Questions 20 and 10 the norms elicited for punisher's contributions 0 and 20, the endpoints, are the same across all types of subjects and across all reference groups. Since in the analysis reported in the main text our arguments rely on the comparisons of average norms (average taken over all potential punisher's contributions), we need to show that the norms are not different at the endpoints. Otherwise, the comparison of average norms might be invalid.

We use Kruskal-Wallis tests to show that the norms for punisher's contributions 0 and 20 are not statistically different. For each of the three questions (Question 20, Question 10, and Punishment Question) we run two sets of tests, one for the punisher's contribution 0 and another for punisher's contribution 20. Since Kruskal-Wallis test assumes independence of the compared groups, we can only compare norms for one reference group for each group of subjects. Thus, we consider the answers in own reference group across good/bad groups and outsiders.

For Question 20 we compare norms in own reference group for punisher's contribution 0 in five groups: good punishers, bad punishers, good victims, bad victims, and outsiders. Kruskal-Wallis test gives p-value of 0.27. Thus, we cannot reject the null hypothesis of equality of distributions of norms for punisher's contribution 0 in own reference group. Similarly, for punisher's contribution 20, Kruskal-Wallis test gives p-value of 0.61. So, for Question 20 and own reference group we can assume that the endpoints of norms are equal, which validates our average norm comparison reported in the main text. Same tests run for Question 10 give insignificant p-values of 0.58 and 0.43 respectively.

We also perform similar tests for the different reference groups. We take answers in own reference group for punishers and victim's answers in punisher reference group.[1] Thus, Kruskal-Wallis tests are run on 4 groups: good punishers, bad punishers, good victims, and bad victims. Similarly, we compare punisher's answers in victims' reference group and victims' answers in own reference group. Eight tests of this kind for both endpoints are insignificant ($p > 0.23$) except one: the test for Question 20, for punisher's contribution 0 when comparing punishers' own reference group and victims' answer in punishers' reference group gives p-value of 0.0228. Performing pair-wise comparisons with ranksum tests, we find that the only group that is significantly different here is bad victims for which the average answer is 1.37 as compared to outsiders groups with averages around 1.1. However, this difference does not invalidate our method of comparing average norms, since it makes the derivative of the norm of bad victims smaller,

[1] We do not include the answers of outsiders here, since in the main text we do not test the differences between outsiders' answers in victim/punisher reference group with those of victims and punishers.

not larger.

To compare endpoints within each group of subjects we cannot use Kruskal-Wallis tests, since the answers to questions related to three reference groups are not independent. Instead we use Friedman test, designed to make such comparisons. We perform 12 Friedman tests, 4 for each group of subjects (punishers, victims, outsiders), out of which 2 are for the two endpoints of Question 20 and 2 for the two endpoints of Question 10. Only two tests out of 12 allow us to reject null hypothesis that the endpoints are the same: one for Question 10 among punishers for endpoint 20 ($p = 0.0053$) and one for Question 10 among outsiders for endpoint 20 ($p = 0.0223$). This, however, does not invalidate our results in the main text, since we do not report significant differences between any groups of subjects for Question 10.

Therefore, overall, we cannot reject the hypotheses that endpoints for norms in Questions 20 and 10 are different for any relevant comparisons and, thus, our method of comparing average norms is valid.

C Variables Used in the Regressions

Variable	Range	Definition
punisher's average contribution	$[0, 20]$	Average contribution of a punisher in 15 rounds of PGG
punisher's total punishment	$[0, 30]$	Sum of punishments of three victims averaged over 15 rounds of PGG
victims's average contribution	$[0, 20]$	Average contribution of a victim in 15 rounds of PGG
xy-q20	$[1, 7]$	Average norm in Question 20 expressed by a subject from group $x \in \{p, v, o\}$ (punishers, victims, outsiders) in a reference group $y \in \{p, v, o\}$ (punishers', victims', outsiders')
xy-q10	$[1, 7]$	Average norm in Question 10 expressed by a subject from group $x \in \{p, v, o\}$ (punishers, victims, outsiders) in a reference group $y \in \{p, v, o\}$ (punishers', victims', outsiders')
xy-qpun	$[1, 7]$	Average norm in Punishment Question expressed by a subject from group $x \in \{p, v, o\}$ (punishers, victims, outsiders) in a reference group $y \in \{p, v, o\}$ (punishers', victims', outsiders')
average norm (own ref. group)	$[1, 7]$	Refers to xx-qz, where $x \in \{p, v, o\}$ and $z \in \{20, 10, \text{pun}\}$, depending on the dependent variable
bad	$0/1$	Is 1 if subject comes from a bad group and 0 if she comes from a good group
punishers	$0/1$	Is 1 for punishers' reference group and 0 for victims' reference group

Table 7: Variables used in the regressions.

D Additional Analyses

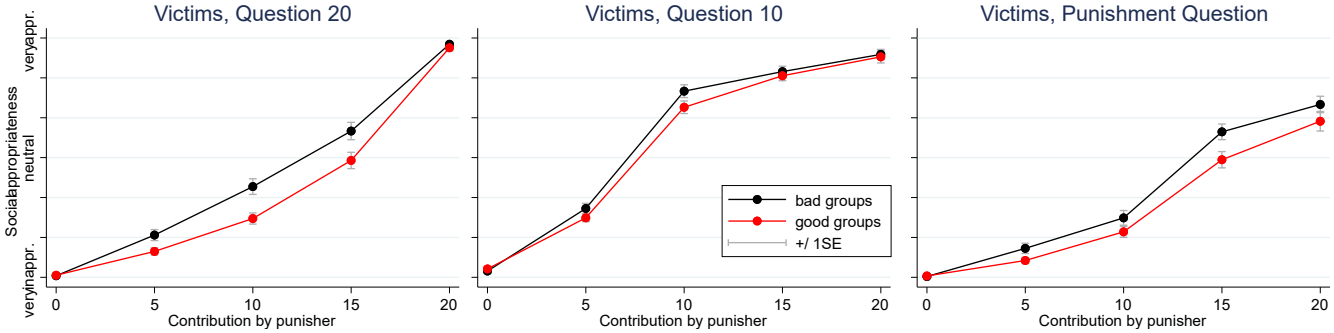

Figure 9: Norms expressed by good and bad victims in their own reference group.

Dependent variable:	pp-q20	pp-q10	pp-qpun
punisher's average contribution	–0.038**	–0.020	–0.029
	(0.016)	(0.013)	(0.023)
constant	3.976***	4.628***	3.513***
	(0.270)	(0.196)	(0.381)
Spearman's rank correlation	–0.319**	–0.332**	–0.257*
N punishers	53	53	53

Table 8: OLS regressions and rank correlations of the norms expressed by punishers on the average punisher's contribution. Errors are robust. Standard errors in parentheses. * – $p < 0.1$; ** – $p < 0.05$; *** – $p < 0.01$.

Dependent variable:	vv-q20	vv-q10	vv-qpun
punisher's average contribution	–0.033***	–0.004	–0.032***
	(0.011)	(0.008)	(0.011)
constant	3.817***	4.441***	3.367***
	(0.202)	(0.138)	(0.164)
Spearman's rank correlation	–0.227***	–0.099	–0.197**
N victims	159	159	159
N groups	53	53	53

Table 9: Random effects regressions and rank correlations of the norms expressed by victims on the average punisher's contribution. Errors are robust. Standard errors in parentheses. * – $p < 0.1$; ** – $p < 0.05$; *** – $p < 0.01$.

	Punishers			Victims		
Dependent variable:	pv-q20	pv-q10	pv-qpun	vp-q20	vp-q10	vp-qpun
average norm (own ref. group)	0.852***	0.490***	0.679***	0.700***	0.546***	0.594***
	(0.075)	(0.096)	(0.177)	(0.074)	(0.096)	(0.099)
constant	0.270	2.187***	0.752	1.209***	2.061***	1.735***
	(0.236)	(0.406)	(0.480)	(0.266)	(0.420)	(0.320)
N observations/subjects	53	53	53	159	159	159
N groups				53	53	53

Table 10: For punishers: OLS regressions of average norms in victims' group. Errors are robust. For victims: random effects regressions of average norms in punishers' group. Errors are clustered by group (of four subjects who play PGG) and robust. ***, **, * denote statistical significance at the 1, 5, and 10 percent level.

	Punishers			Victims		
Dependent variable:	po-q20	po-q10	po-qpun	vo-q20	vo-q10	vo-qpun
average norm (own ref. group)	0.977***	0.546***	0.683***	0.722***	0.512***	0.682***
	(0.078)	(0.106)	(0.127)	(0.060)	(0.088)	(0.069)
constant	–0.159	1.917***	1.033**	0.823***	2.035***	1.126***
	(0.245)	(0.465)	(0.389)	(0.198)	(0.402)	(0.218)
N observations/subjects	53	53	53	159	159	159
N groups				53	53	53

Table 11: For punishers: OLS regressions of average norms in outsiders' group. Errors are robust. For victims: random effects regressions of average norms in outsiders' group. Errors are clustered by group (of four subjects who play PGG) and robust. ***, **, * denote statistical significance at the 1, 5, and 10 percent level.

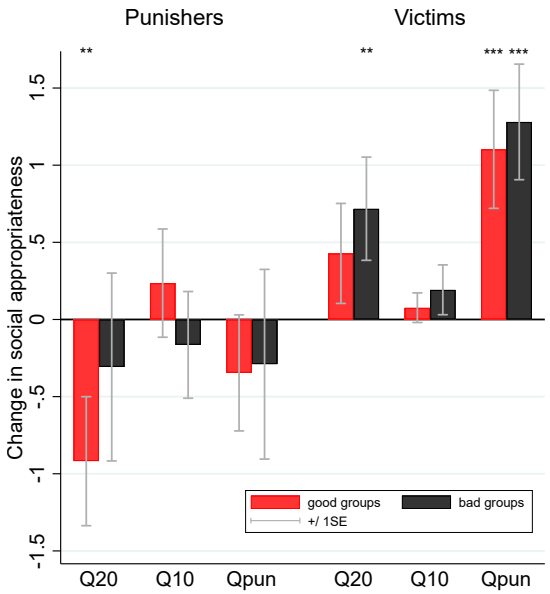

Figure 10: Estimates of g_τ minus the average norms in own reference group for punishers and victims. Above zero values mean that punishers/victims think that victims/punishers consider actions in a given question more socially appropriate than they themselves do in their own reference group. The significance levels reported are those of the non-linear transformations of the coefficients from the regressions in Table 12 minus average norm in own reference group. ***, **, * - significance at the 1%, 5%, and 10%.

	Punishers			Victims		
Dependent variable:	pv-q20	pv-q10	pv-qpun	vp-q20	vp-q10	vp-qpun
bad	0.259	–0.001	0.168	0.182*	0.146	0.216
	(0.182)	(0.192)	(0.257)	(0.109)	(0.097)	(0.158)
average norm (own ref. group)	0.771***	0.643***	0.635***	0.738***	0.450***	0.593***
	(0.105)	(0.190)	(0.220)	(0.076)	(0.108)	(0.123)
constant	0.523	1.544**	0.907	0.944***	2.414***	1.545***
	(0.310)	(0.720)	(0.615)	(0.255)	(0.472)	(0.387)
N observations/subjects	36	36	36	108	108	108
N groups				36	36	36

Table 12: For punishers: OLS regressions of average norms in victims' group. Errors are robust. For victims: random effects regressions of average norms in punishers' group. Errors are clustered by group (of four subjects who play PGG) and robust. ***, **, * denote statistical significance at the 1, 5, and 10 percent level.

Dependent variable:	ov-q20/ op-q20	ov-q10/ op-q10	ov-qpun/ op-qpun
punishers	0.420***	0.169*	0.739***
	(0.119)	(0.098)	(0.107)
average norm (own ref. group)	0.539***	0.456***	0.811***
	(0.087)	(0.144)	(0.072)
constant	1.183***	2.331***	0.271
	(0.303)	(0.663)	(0.206)
N observations	118	118	118
N subjects	59	59	59

Table 13: Outsiders: OLS regressions of average norms in victims' and punishers' groups. Errors are robust. ***, **, * denote statistical significance at the 1, 5, and 10 percent level.

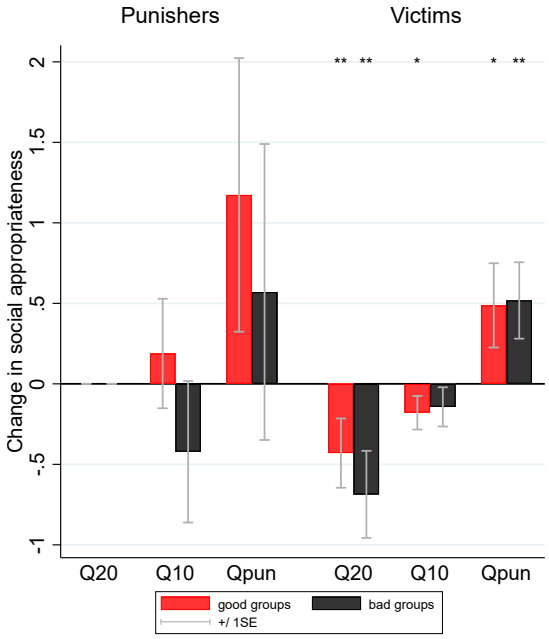

Figure 11: Estimates of g_τ in outsiders' reference group minus the average norms in own reference group for punishers and victims. The significance levels are those of the non-linear transformations of the coefficients from the regressions in Table 14 minus average norm in own reference group. ***, **, * denote statistical significance at the 1, 5, and 10 percent level.

	Punishers			Victims		
Dependent variable:	po-q20	po-q10	po-qpun	vo-q20	vo-q10	vo-qpun
bad	0.038	-0.059	-0.041	0.034	0.102	0.113
	(0.162)	(0.148)	(0.194)	(0.085)	(0.080)	(0.099)
average norm (own ref. group)	0.996***	0.721***	0.794***	0.772***	0.452***	0.705***
	(0.093)	(0.152)	(0.134)	(0.061)	(0.096)	(0.080)
constant	-0.178	1.192*	0.826*	0.626***	2.262***	0.938***
	(0.296)	(0.643)	(0.450)	(0.177)	(0.434)	(0.240)
N observations/subjects	36	36	36	108	108	108
N groups				36	36	36

Table 14: For punishers: OLS regressions of average norms in others' reference group. Errors are robust. For victims: random effects regressions of average norms in others' reference group. Errors are clustered by group (of four subjects who play PGG) and robust. ***, **, * denote statistical significance at the 1, 5, and 10 percent level.

E Instructions

E.1 Public Goods Game Instructions

General information
You are about to participate in a decision making experiment. If you follow the instructions carefully, you can earn a considerable amount of money depending on your decisions and the decisions of the other participants. Your earnings will be paid to you in cash at the end of the experiment.

This set of instructions is for your private use only. During the experiment, you are not allowed to communicate with anybody. In case of questions, please raise your hand. Then we will come to your seat and answer your questions. Any violation of this rule excludes you immediately from the experiment and all payments. The funds for conducting this experiment were provided by Max Planck Institute for Research on Collective Goods.

Throughout the experiment, you will make decisions about amounts of tokens. At the end of the experiment, all tokens you have will be converted into Euros at the exchange rate 0.20 Euro for 1 token and paid you in cash in addition to the show-up fee of 5 Euros.

During the experiment, all your decisions will be treated confidentially. This means that none of the other participants will be able to associate your decisions with your personal identity.

PART I

Part I of the experiment will consist of 15 decision making periods. At the beginning of the experiment, you will be matched with 3 other people in this room. Therefore, there are 4 people, including yourself, participating in your group. You will be matched with the same people during the entire Part I of the experiment. For the purpose of the experiment, you and the other group members will be randomly assigned labels A, B, C, and D that will identify you and others throughout the Part I of the experiment. None of the participants knows your personal identity in the group.

First Stage of a Period
Before each period you, and each other person in your group, will be given the endowment of 20 tokens. At the first stage of each period, you will be asked to allocate your endowment between a private account and a group account. Other members of your group will be asked to do the same. The tokens that you place in the private account have a return of 1. This means that at the end of the first stage of each period your private account will contain exactly the amount of tokens you put into the private account at the beginning of the period. Nobody except yourself benefits from your private account. The tokens that you place in the group account are summed together with the tokens that the other three members of your group place in the group account. The tokens in the group account have a return of 2. Every member of the group benefits equally from the group account. Specifically, the total amount of tokens placed in the group account by all group members is multiplied by 2 and then is equally divided among the four group members. Hence, your share of the group account is

$$2 * (\text{sum of tokens in the group account})/4$$

Thus, at the end of the first stage of each period, the number of tokens that you have is equal to the number of tokens you place in your private account plus your share of the group account.

$$\text{Payoff} = 20 - \text{tokens you put into the group account} + 2 * (\text{sum of tokens in the group account})/4$$

Here are three examples to make this clear:

1. Suppose you place 0 tokens in the group account and 20 tokens in the private account, and the other members of your group place a total of 45 tokens in the group account. The sum of tokens in the group account is 45. Your share of the group account would be 2 * 45 / 4 = 22.5 tokens. Each other

member of the group would also receive a share of the group account equal to 22.5 tokens. The amount of tokens that you have at the end of the first stage is, thus, equal to 20 + 22.5 = 42.5 tokens. Each other member of your group receives on average 27.5 tokens.

2. Suppose you place 15 tokens in the group account and 5 tokens in the private account, and the other members of your group place a total of 45 tokens in the group account. The sum of tokens in the group account is 60. Your share of the group account would be 2 * 60 / 4 = 30 tokens. Each other member of the group would also receive a share of the group account equal to 30 tokens. The amount of tokens that you have at the end of the first stage is, thus, equal to 5 + 30 = 35 tokens. Each other member of your group receives on average 35 tokens.

3. Suppose you place 15 tokens in the group account and 5 tokens in the private account, and the other members of your group place a total of 10 tokens in the group account. The sum of tokens in the groups account is 25. Your share of the group account would be 2 * 25 / 4 = 12.5 tokens. Each other member of the group would also receive a share of the group account equal to 12.5 tokens. The amount of tokens that you have at the end of the first stage is, thus, equal to 5 + 12.5 = 17.5 tokens. Each other member of your group receives on average 29.1 tokens.

Second Stage of a Period
In the second stage of each period, only the member of your group who was labeled D is active. The group members who received labels A, B, and C do not make any decisions in the second stage of each period.

If your label in the group is D, you will be asked to react to the decisions made by group members A, B, and C during the first stage of each period. At this point, you will already know the decisions taken by each group member at the first stage and the number of tokens they have after the first stage. You will decide whether you want to subtract tokens from any other group member or not. The group members that you decide to subtract tokens from will lose the amount of tokens you choose. The decisions you make at this stage will not change the amount of tokens that you have after the first stage.

You may subtract different amounts of tokens from different group members. The total amount of tokens that you choose to subtract from the group members A, B, and C may not exceed 30 tokens. Any group member can only lose maximum the amount of tokens he or she has. For example, if at the end of the first stage group members A, B, and C have 10, 15, and 20 tokens respectively, and you choose to subtract 15, 10, and 0 tokens from them then group members A, B, and C will be left with 0, 5, and 20 tokens.

Information about the Choices and Tokens in the End of a Period
At the end of each period, each member of the group will be informed about:

- His/her contribution to the group account;
- The amount of tokens contributed by all group members individually to the group account;
- His/her share of the group account (remember, it is the same for all group members);
- If you are member A, B, or C: how many tokens were subtracted from you by member D;
- If you are member A, B, or C: the number of tokens at the end of the period, which is equal to the number of tokens in the private account plus the share of tokens from the group account minus the number of tokens subtracted by D;
- If you are member D: the number of tokens at the end of the period, which is equal to the number of tokens in the private account plus the share of tokens from the group account.

Structure of Part I of the Experiment
The structure of the experiment in all 15 periods is identical. In the first stage of each period each group member A, B, C, and D chooses how to split 20 tokens between private and group accounts. Then all group members receive the returns from both accounts. In the second stage of the period, group member D can subtract tokens from group members A, B, and C. In the end of the period all members are informed about the decisions of others in the group, and the number of tokens they have.

Money Earned in Part I of the Experiment
In the end of the experiment, the computer will randomly choose one period for which you and other members of your group will be paid. Your income at the end of Part I of the experiment is equal to the amount of tokens at the end of this randomly chosen period times the exchange rate of 0.20 Euro for 1 token.

This is the end of the instructions for Part I. If you have any questions please raise your hand and an experimenter will come by to answer them.

E.2 Norm Elicitation Instructions for PGG subjects

PART II

Description of the Task (Screen 1)
On the following screens, you will read the descriptions of a series of hypothetical situations that could have taken place in Part I of the experiment. These descriptions correspond to situations in which a person, acting in the role of member D (who will be called Individual D), makes decisions about the amounts of tokens to be placed in the group account and decisions to subtract tokens from members A, B, and C. For each situation, you will be given a description of the decision faced by Individual D. This description will include several possible choices available to this Individual.

After you read the description of the decision, you will be asked to evaluate the different possible actions available and to decide, for each of the actions, whether taking that action would be ”socially appropriate” and ”consistent with moral or proper social behavior” or ”socially inappropriate” and ”inconsistent with moral or proper social behavior.” By socially appropriate, we mean behavior that most people agree is the ”correct” or ”ethical” thing to do. Another way to think about what we mean is that if Individual D were to select a socially inappropriate choice, then someone else might be angry at Individual D for doing so.

In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behavior.

To give you an idea of how the experiment will proceed, we will go through an example and show you how you will indicate your responses. On the next screen you will see an example of a situation. Click OK when you are ready to go on.

Example Situation (Screen 2)
Bob is at a café. While there, Bob notices that someone has left a wallet at one of the tables. Bob must decide what to do. He has four possible choices: take the wallet, ask others nearby if the wallet belongs to them, leave the wallet where it is, or give the wallet to the bartender. Bob can choose only one of these four options. The table on the right of the screen presents a list of the possible actions available to Bob. For each of the actions, please indicate on the scale from 1 to 7 how socially appropriate you believe choosing that option is. To indicate your response, please click on the corresponding cell. Please make sure you make an assessment for each possible choice in each row of the table.

Screen 3
In what follows, you will be asked to assess the appropriateness of the actions in three situations that

could have arisen in Part I of the experiment. For each action in each situation please indicate the extent to which you believe taking that action would be "socially appropriate" and "consistent with moral or proper social behavior" or "socially inappropriate" and "inconsistent with moral or proper social behavior." By socially appropriate we mean behavior that most people agree is the "correct" or "ethical" thing to do.

Payment
For each situation that follows, you will read its description. You will then indicate your appropriateness rating by placing a check mark in the corresponding cell.

At the end of Part II of the experiment, in order to determine your payment, we will randomly select one of the situations. For this situation, we will also randomly select one of the possible choices that Individual D could make. Thus, we will select both a scenario and one possible choice at random. This means that when you make your choices you should make each of them as if it is the one for which you will be paid.

Your payment in this part of the experiment will depend on whether your response to the choice thus selected is the same as the response made by the most people with the same role as you in Part I of the experiment (who are in this room). In particular, if in Part I of the experiment you were member A, B, or C then your response to a selected choice will be compared to the responses of all people in this room who were members A, B, and C in Part I. If you were member D, then your response to a selected choice will be compared to the responses of all people in this room who were members D. If you give the same response as that most frequently given by other members with the same role, then you will receive €8. This amount will be paid to you, in cash, at the conclusion of the experiment.

For instance, there are overall $N/4$ participants who were members D in the previous part of the experiment and $3N/4$ participants who were members A, B, or C (including you). Suppose we were to select the example situation from the last screen and the possible choice "Leave the wallet where it is," and your response had been 3, "somewhat socially inappropriate." Then, if you are member D, you would receive €8 if this was the response selected by most of other $N/4-1$ members D in today's session. If you are member A, B, or C, you would receive €8 if this was the response selected by most of other $3N/4-1$ members A, B, and C in today's session. If your response is not the same as that of the majority of others with the same role as you, you will receive nothing in this part of the experiment.

Please click OK when you are ready to go on. If you have any questions, please raise your hand and wait for the experimenter to come.

Screen 4
Imagine that members A, B, C have each placed 10 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by other ⟨NUMBER⟩ members ⟨ROLE⟩ in this room.

Screen 5
Imagine that members A, B, C have each placed 20 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by other ⟨NUMBER⟩ members ⟨ROLE⟩ in this

room.

Screen 6
Imagine that members A, B, C, and D have made their choices in the first stage of a period. Namely, members A, B, and C placed 10 tokens each to the group account and individual D placed the amount of tokens equal to one of the five options listed on the right part of the screen. For each of the amounts that individual D could have placed to the group account, please indicate how socially appropriate you believe subtracting tokens from individuals A, B, and C is, given the amount that members A, B, C, and D contributed to the group account.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by other ⟨NUMBER⟩ members ⟨ROLE⟩ in this room.

PART III

Description of the Task (Screen 1)
In this final part of the experiment we ask you to evaluate the social appropriateness of actions in the same three situations as before. The only difference is that now you will be paid if your evaluation is the same as the evaluation of the majority of two groups of participants who have already made their evaluation decisions. The first group is the participants who had other role than you (members ⟨OTHER ROLE⟩ in this room) who have just made their evaluations in Part II. The second group is a separate group of other participants who took part in the experiment before and who evaluated the same situations as in the previous part but without actually making real choices as in Part I. In particular, these other participants were given the same instructions of Part I as you did and then evaluated social appropriateness in exactly same way that you just did, with the only difference that for the payment they were matched with everyone in their respective sessions.

Payment (Screen 2)
As before, for your payment we will choose one random situation and one random action that you evaluate. This means that when you make your choices you should make each of them as if it is the one for which you will be paid. Your payment in this part of the experiment will depend on whether your response to the selected choice is the same as the response made by the most people in a group who have already chosen. For example, if you are matched with members ⟨OTHER ROLE⟩, then your payment depends on how members ⟨OTHER ROLE⟩ chose in the previous part of the experiment. Remember, the members ⟨OTHER ROLE⟩ when choosing in Part II were paid if they chose the same answer as the majority of other members ⟨OTHER ROLE⟩. The same holds for the separate group of other participants. If you are matched with them, then your payment depends on how they chose in a separate experiment. Remember, these participants were paid if they chose the same answer as the majority of other participants in their session.

If you give the same response as that most frequently given by other members in one of the two groups, then you will receive €8. This amount will be paid to you, in cash, at the conclusion of the experiment. Please click OK when you are ready to go on. If you have any questions, please raise your hand and wait for the experimenter to come.

Screen 4
Put yourself in the shoes of MEMBERS ⟨OTHER ROLE⟩ in this room who have just provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority in their own group of members ⟨OTHER ROLE⟩. Imagine that members A, B, C have each placed 10 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the

amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by MEMBERS ⟨OTHER ROLE⟩ in this room in the previous part of the experiment.

Screen 5

Put yourself in the shoes of MEMBERS ⟨OTHER ROLE⟩ in this room who have just provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority in their own group of members ⟨OTHER ROLE⟩.

Imagine that members A, B, C have each placed 20 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by MEMBERS ⟨OTHER ROLE⟩ in this room in the previous part of the experiment.

Screen 6

Put yourself in the shoes of MEMBERS ⟨OTHER ROLE⟩ in this room who have just provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority in their own group of members ⟨OTHER ROLE⟩.

Imagine that members A, B, C, and D made their choices in the first stage of a period. Namely, members A, B, and C placed 10 tokens each to the group account and individual D placed the amount of tokens equal to one of the five options listed on the right part of the screen. For each of the amounts that individual D could have placed to the group account, please indicate how socially appropriate you believe subtracting tokens from individuals A, B, and C is, given the amount that they contributed to the group account.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by MEMBERS ⟨OTHER ROLE⟩ in this room in the previous part of the experiment.

Screen 7

Put yourself in the shoes of OTHER PARTICIPANTS who gave evaluations in the previous experiment who have provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority in their own group.

Imagine that members A, B, C have each placed 10 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by OTHER PARTICIPANTS in a separate the experiment.

Screen 8

Put yourself in the shoes of OTHER PARTICIPANTS who gave evaluations in the previous experiment who have provided their evaluations of social appropriateness of the actions of Individual D in the fol-

lowing situation that you have also seen. Remember, that they were paid if they guessed as the majority in their own group.

Imagine that members A, B, C have each placed 20 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by OTHER PARTICIPANTS in a separate the experiment.

Screen 9

Put yourself in the shoes of OTHER PARTICIPANTS who gave evaluations in the previous experiment who have provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority in their own group.

Imagine that members A, B, C, and D made their choices in the first stage of a period. Namely, members A, B, and C placed 10 tokens each to the group account and individual D placed the amount of tokens equal to one of the five options listed on the right part of the screen. For each of the amounts that individual D could have placed to the group account, please indicate how socially appropriate you believe subtracting tokens from individuals A, B, and C is, given the amount that they contributed to the group account.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by OTHER PARTICIPANTS in a separate the experiment.

E.3 Instructions for Outsiders

PART I

Description of the Experiment (Screen 1)

On the following screens, you will read the descriptions of a series of hypothetical situations. These descriptions correspond to situations in which one person, Individual D, must make a decision. For each situation, you will be given a description of the decision faced by Individual D. This description will include several possible choices available to this Individual.

After you read the description of the decision, you will be asked to evaluate the actions available to Individual D and to decide, for each of the actions, whether taking that action would be "socially appropriate" and "consistent with moral or proper social behavior" or "socially inappropriate" and "inconsistent with moral or proper social behavior." By socially appropriate, we mean behavior that most people agree is the "correct" or "ethical" thing to do. Another way to think about what we mean is that if Individual D were to select a socially inappropriate option, then someone else might be angry at Individual D for doing so.

In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behavior.

To give you an idea of how the experiment will proceed, we will go through an example and show you how you will indicate your responses. On the next screen you will see an example of a situation. Click OK when you are ready to go on.

Example Situation (Screen 2)

Bob is at a café. While there, Bob notices that someone has left a wallet at one of the tables. Bob must decide what to do. He has four possible choices: take the wallet, ask others nearby if the wallet belongs

to them, leave the wallet where it is, or give the wallet to the bartender. Bob can choose only one of these four options. The table on the right of the screen presents a list of the possible actions available to Bob (in rows). For each of the actions, please indicate on the scale from 1 to 7 how socially appropriate you believe choosing that option is. To indicate your response, please click on the corresponding cell.

Please make sure you make an assessment for each possible choice in each row of the table.

Screen 3

In what follows, you will be asked to assess the appropriateness of the actions in three situations similar to the one you have just seen. For each action in each situation please indicate the extent to which you believe taking that action would be "socially appropriate" and "consistent with moral or proper social behavior" or "socially inappropriate" and "inconsistent with moral or proper social behavior." By socially appropriate we mean behavior that most people agree is the "correct" or "ethical" thing to do.

Payment

For each situation that follows, you will read its description. You will then indicate your appropriateness rating by placing a check mark in the corresponding cell.

At the end of the experiment, in order to determine your payment, we will randomly select one of the situations. For this situation, we will also randomly select one of the possible choices that Individual D could make. Thus, we will select both a scenario and one possible choice at random. This means that when you make your choices you should make each of them as if it is the one for which you will be paid.

Your payment in this part of the experiment will depend on whether your response to the choice thus selected is the same as the response made by the most people in this room.

If you give the same response as that most frequently given by other participants, then you will receive €8. This amount will be paid to you, in cash, at the conclusion of the experiment.

For instance, if we were to select the example situation from the last screen and the possible choice "Leave the wallet where it is," and if your response had been 3, "somewhat socially inappropriate," then you would receive €8, in addition to the €5 participation fee, if this was the response selected by most other people in today's session. Otherwise you would receive only the €5 participation fee.

Please click OK when you are ready to go on. If you have any questions, please raise your hand and wait for the experimenter to come.

Description of the Situation (Screen 4 and print-out)

Individual D has been invited to an experiment and placed in a group with three other anonymous people labeled A, B, and C so that no individual will ever know the identity of the other individuals with whom he/she is grouped. In fact, suppose that individuals A, B, C, and D are part of a larger group of people participating in this experiment, exactly as you are now. Individuals A, B, C, and D are given experimental instructions exactly as those you can find on your desk.

In order to understand what decisions Individual D has to make, please read these instructions carefully.

On the following screens you will be asked to evaluate social appropriateness of the actions of Individual D. Each screen will show the description of choices made by individuals A, B, and C and you will be asked to guess how socially appropriate several actions of individual D are.

Please click OK when you have read the instructions and are ready to go on.

Screen 5

Imagine that individuals A, B, C have each placed 10 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a situation and an action for payment, you will only receive €8 if your

response is the same as the most frequent response made by other participants in this room.

Screen 6
Imagine that individuals A, B, C have each placed 20 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a situation and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by other participants in this room.

Screen 7
Imagine that individuals A, B, C, and D made their choices in the first stage of a period. Namely, individuals A, B, and C placed 10 tokens each to the group account and individual D placed the amount of tokens equal to one of the five options listed on the right part of the screen. For each of the amounts that individual D could have placed to the group account, please indicate how socially appropriate you believe subtracting tokens from individuals A, B, and C is, given the amount that Individuals A, B, C, and D contributed to the group account.

Remember: when we select a situation and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by other participants in this room.

PART II

Description of the Task (Screen 1)
In this part of the experiment we ask you to evaluate the social appropriateness of actions in the same three situations as before. The only difference is that now you will be paid if your evaluation is the same as the evaluation of the majority of two distinct groups of participants who have already made their evaluation decisions in a previous experiment.

In the previous sessions that we ran in this lab we had participants who have actually made choices in the experiment described in the instructions on your desk. After that these participants evaluated the appropriateness of the same situations that you have just seen and were paid if their guesses were the same as those given by the majority of participants in the same role. To understand how exactly this was happening, imagine that you are individual D who has just made choices in the experiment described in the instructions on your desk (which has actually happened in previous sessions). After that you are asked to evaluate the appropriateness of the same situations that you have seen in the previous part of the experiment, but you are told that you will be paid only if your evaluation of a randomly chosen action in one of the three situations is the same as the evaluation of the majority of other participants in the role of individual D in the session. Or similarly, imagine that you are individual A, B, or C and you have just made choice in the experiment. Then you are asked to provide evaluations of appropriateness of actions of individual D and you are paid if the majority of other participants in the role of individuals A, B, and C in the session gave the same answers.

To summarize, in what follows you will be asked to evaluate social appropriateness of the same actions in the same situations you have already seen, but your payment will depend on the answers of participants in two distinct groups: 1) participants who actually chose in the experiment as individuals D and were later matched with other individuals D for appropriateness evaluations and 2) participants who actually chose in the experiment as individuals A, B, and C and were later matched with other individuals A, B, and C for appropriateness evaluations.

Payment (Screen 2)
As before, for your payment we will choose one random situation and one random action that you evaluate. This means that when you make your choices you should make each of them as if it is the one

for which you will be paid. Your payment in this part of the experiment will depend on whether your response to the selected choice is the same as the response made by the most people in one of the two groups as described on the previous screen. For example, if you are matched with individuals D from previous experiment, then your payment depends on how these individuals evaluated the appropriateness of the same actions when matched with other individuals D in their session. The same holds when you are matched with individuals A, B, and C.

If you give the same response as that most frequently given by other members in one of the two groups, then you will receive €8. This amount will be paid to you, in cash, at the conclusion of the experiment.

Please click OK when you are ready to go on. If you have any questions, please raise your hand and wait for the experimenter to come.

Screen 3

Put yourself in the shoes of INDIVIDUALS D who took part in a previous experiment and have provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority of other individuals D in their own group.

Imagine that members A, B, C have each placed 10 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by INDIVIDUALS D in a separate the experiment.

Screen 4

Put yourself in the shoes of INDIVIDUALS D who took part in a previous experiment and have provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority of other individuals D in their own group.

Imagine that members A, B, C have each placed 20 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by INDIVIDUALS D in a separate the experiment.

Screen 5

Put yourself in the shoes of INDIVIDUALS D who took part in a previous experiment and have provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority of other individuals D in their own group.

Imagine that members A, B, C, and D made their choices in the first stage of a period. Namely, members A, B, and C placed 10 tokens each to the group account and individual D placed the amount of tokens equal to one of the five options listed on the right part of the screen. For each of the amounts that individual D could have placed to the group account, please indicate how socially appropriate you believe subtracting tokens from individuals A, B, and C is, given the amount that they contributed to the group account.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by INDIVIDUALS D in a separate the experiment.

Screen 6

Put yourself in the shoes of INDIVIDUALS A, B, and C who took part in a previous experiment and have provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority of other individuals A, B, and C in their own group.

Imagine that members A, B, C have each placed 10 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by INDIVIDUALS A, B, and C in a separate the experiment.

Screen 7

Put yourself in the shoes of INDIVIDUALS A, B, and C who took part in a previous experiment and have provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority of other individuals A, B, and C in their own group.

Imagine that members A, B, C have each placed 20 tokens (out of 20) to the group account in the previous period. Look at the table on the right side of the screen and consider five possible amounts that Individual D could place in the group account (presented in rows). Please indicate on the scale from 1 to 7 how socially appropriate you believe choosing each of these amounts is, given the amounts that others contributed to the group account in the previous period.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by INDIVIDUALS A, B, and C in a separate the experiment.

Screen 8

Put yourself in the shoes of INDIVIDUALS A, B, and C who took part in a previous experiment and have provided their evaluations of social appropriateness of the actions of Individual D in the following situation that you have also seen. Remember, that they were paid if they guessed as the majority of other individuals A, B, and C in their own group.

Imagine that members A, B, C, and D made their choices in the first stage of a period. Namely, members A, B, and C placed 10 tokens each to the group account and individual D placed the amount of tokens equal to one of the five options listed on the right part of the screen. For each of the amounts that individual D could have placed to the group account, please indicate how socially appropriate you believe subtracting tokens from individuals A, B, and C is, given the amount that they contributed to the group account.

Remember: when we select a scenario and an action for payment, you will only receive €8 if your response is the same as the most frequent response made by INDIVIDUALS A, B, and C in a separate the experiment.