ORIGINAL ARTICLE OPEN ACCESS

Normative Acceptance of Power Abuse

Leonard Hoeft[1] | Wladislaw Mill[2] (ID) | Alexander Vostroknutov[3]

[1]Department of Law, Humboldt University Berlin, Berlin, Germany | [2]Department of Economics, University of Mannheim, Mannheim, Germany | [3]School of Business and Economics, Maastricht University, Maastricht, Netherlands

Correspondence: Wladislaw Mill (mill@uni-mannheim.de)

ABSTRACT

While extensive research has shown how power corrupts those who wield it, less is known about its effects on the powerless. Victims of power abuse may either become dissatisfied with such norm violations by those in power or normalize such behavior, corrupting their own normative standards. In a laboratory experiment, we create communities where a single individual holds the power to punish others. We find that there is general agreement that it is socially inappropriate for those in power both to undercontribute and to enforce higher contributions from others. Yet we find that the powerless who are randomly exposed to this behavior believe it to be less socially inappropriate than those who are spared from it. These findings shed light on how power abuse can perpetuate itself and contribute to the persistence of corrupt institutions.

JEL Classification: C91, C92, K42, H41, D73

1 | Introduction

Throughout human history, societies always faced a dilemma: they typically rely on institutions to maintain cooperation and prosocial behavior, but institutions also create opportunities for abuse of power (Acemoglu and Robinson 2019; Robinson et al. 2023). Conferring power to individuals or groups might induce selfish behavior and norm violations, as is evidenced by the long line of research on the corrupting effects of power (e.g., Fehr et al. 2013; Falk and Kosfeld 2006; Hoeft and Mill 2017). In particular, people in the position of power may gain an unfair advantage by not being subject to norm enforcement themselves (i.e., they "play a rigged game," Acemoglu and Robinson 2008; Dal Bó et al. 2009); politicians may use their influence to attain atypical benefits (Grant and Keohane 2005; Olken and Pande 2012); police officers may use illegal violence (Wong 1998); doctors may use their connections for special treatment (Klitzman 2007, pp. 71–73); managers may force their coworkers to invest in shared projects that they themselves skimp on (Xu et al. 2015;

Vredenburgh and Brender 1998). All these are examples of abuse of power. The list of instances where people elevated to power misuse it is long and the effect is well-established. However, one commonly ignored aspect of this social dynamics is how *victims* of power abuse judge the abuse of power.

One could expect that victims of power abuse become more aware it, as they personally bear the costs of such norm violations, and judge such acts as immoral and/or harmful to the community. In such cases—provided that existing institutions allow it—public dissatisfaction may result in actions towards reducing the harmful effects of power abuse (e.g., the ruling political party does not get re-elected). However, victims may also justify or exonerate abusive behavior to reduce the cognitive dissonance between their own experience and the world (Furnham 2003). In this case, power abuse may persist unchecked for some time. Which of the two effects is larger can have significant consequences for policies aimed at battling power abuse. If power abuse has an indirect, corrupting effect

on victims, simply replacing those in power may not be enough, as the previously powerless may have been corrupted themselves and fail to recognize the negative effects of abuse.

Understanding how victims of power abuse perceive and normatively judge its effects is practically impossible with observational data. Specifically, causal evidence for the hypotheses like above is hard to establish in the real world as the distribution of power, the prevailing norms, and the selection of the powerful are all endogenous to each society. Therefore, we turn to the lab and create artificial mini-societies with power asymmetries to provide a causal answer. We implement a repeated Public Goods game, where only one randomly chosen member, whom we call *punisher,* has the power to punish others at no personal cost, thereby representing a designated sanctioning institution. Through punishment, this member can dictate contribution norms without personally adhering to them (Hoeft and Mill 2017, 2024), which constitutes a combination of undercontribution and "hypocritical" punishment. As we find in our present experiments, both of these are unanimously considered a violation of social norms among the powerful, the powerless, and the third parties. To establish causal effect in our main research question—that is, how the experience of abusive behavior affects victims—we use the fact that the powerless are randomly assigned to the powerful and thereby to those who use their power inappropriately and those who do not. This randomization allows us to compare group averages of victims' attitudes after experiencing abuse or not and to make a causal claim that the presence of the abuse leads to the observed differences.

To establish if and how experiencing abusive behavior affects victims' attitudes, we use a well-known incentive-compatible mechanism (Krupka and Weber 2013) to elicit *social norms* among the powerless regarding the behavior of the powerful in different scenarios (social norms are beliefs about what the majority in some reference group thinks is appropriate). Our null hypothesis is that usual participants, or *nonpunishers*, rate the social appropriateness of abusive behavior in the same way regardless of their experience. The alternative hypothesis is that power, once abused, can even corrupt the nonpunishers as well.

We document for the first time that *nonpunishers* who are randomly assigned to an "abusive" punisher consider it significantly *more* appropriate for the punishers to abuse their power than participants who are randomly assigned to a "nonabusive" punisher. Specifically, the comparison of nonpunishers who experienced abuse and those who did not shows that the former group thinks that *from the societal perspective* the punishers' free-riding is more socially appropriate than the latter group. Similarly, nonpunishers with experience of abuse indicate that punishing others while contributing less is socially *more* appropriate than what nonpunishers without abuse experience report. These effects become even stronger as participants experience abuse for more periods.

Our findings contribute to several strands of literature. First, we contribute to the literature documenting the corruption effect of power, which revealed that power leads to selfishness (Kipnis 1972; Galinsky et al. 2015; Lammers et al. 2015), cheating (Lammers et al. 2010; Dubois et al. 2015), and distrust

(Schilke et al. 2015). People in power also violate social norms (van Kleef et al. 2015a), care less about the payoffs of others (Guinote et al. 2015), and impose unjust allocations in economic games (Giurge et al. 2021; van Dijk et al. 2004; Bendahan et al. 2015; Werner et al. 2024). Importantly, elevating people to a position of power modifies their personal values (Rustichini and Villeval 2014; Mallucci et al. 2019; Schier et al. 2016) as they become more strict in judging and punishing others (Lammers et al. 2010; Wiltermuth and Flynn 2013; Mooijman et al. 2015). We add to this literature by studying the potential corrupting effects of power not on the powerful themselves, but on the powerless. We demonstrate that even those who bear the cost of norm violations are more accepting of this behavior if they experienced it before. This also marks a striking difference to our previous work (Hoeft and Mill 2024, 2017), which established that power leads to selfishness, but did so only for the powerful. The present study offers a new perspective by documenting the corrupting effects of power on the powerless.

Second, we contribute to the literature on normative judgment and social learning. It has been widely documented that witnessing the behavior of others can erode norm compliance through the repetition of the same behavior (Lindström et al. 2018; Bicchieri et al. 2022; Fischbacher et al. 2001; Merguei et al. 2022; Bicchieri and Xiao 2009; Hoeft et al. 2025). Such spillovers take place in situations where all agents can choose the same actions (and thus they copy each other). We, however, show that observing the actions of others *that cannot be chosen by the agents themselves* can shift normative expectations. While this relates to recent evidence that observing unfair behavior might foster unfair behavior (Herz and Taubinsky 2017; Banerjee 2016; Gächter and Schulz 2016), we provide evidence that experiencing unfair behavior can lead to the normative acceptance of this behavior, speaking to a more fundamental psychological mechanism behind the findings in the literature.

Third, we contribute to the literature on the contagiousness of norm violations. By documenting that even the victims of norm violations start to normatively accept this behavior, we provide a possible mechanism for why criminal behavior is often spatially correlated (Glaeser et al. 1996; Zenou 2003), why diplomats from corrupt countries commit more parking violations (Fisman and Miguel 2007), and why normative values in regions that experienced the rule of despotic institutions are different from countries that did not endure such rule (Tabellini 2008, 2010).

Fourth, we contribute to a broader psychological and sociological literature on rule acceptance and compliance (see, e.g., the overview articles Cialdini and Goldstein 2004; Tyler 2006). A substantial body of research suggests that when rules are enforced in particular ways, the very behaviors they sanction may not only be accepted but even internalized as normative. Specifically, individuals exposed to well-constructed rules often come to view actions aligned with these rules as inherently appropriate, effectively normalizing the behavior (Cialdini et al. 1990; Tyler and Lind 1992; Ashforth and Anand 2003). For example, studies on social influence and conformity—ranging from classic experiments (e.g., Asch 1956) to more recent analyses of behavior (e.g., Cialdini 2013)—demonstrate that repeated exposure to sanctioned behavior can gradually shift individuals' normative

evaluations. Moreover, research by Tyler (2000) and Tyler and Blader (2003) has shown that compliance is fostered when rules are seen as just and institutions are viewed as legitimate. More recent work highlights the dynamic nature of norm change and erosion (Gelfand et al. 2024; Alger et al. 2024; Bicchieri and Garzino Demo 2025). For example, recent studies show that interpersonal reactions to norm violations (van Kleef et al. 2015b) and exposure to deviant behaviors (Álvarez-Benjumea, 2022) can accelerate normative shifts, while pluralistic ignorance can sustain even harmful norms (Smerdon et al. 2020). We contribute to this broader literature by providing experimental evidence that even those suffering from abusive behavior shift their normative evaluations of these behaviors to view it as legitimate.

All in all, our paper provides new evidence that power abuse might also corrupt the normative views of its victims. We document that participants, who experienced abusive behavior, report normative ratings that are more accepting of such behavior. These insights draw a rather grim picture in which the powerful abuse their position, believing that they have done nothing wrong, while the powerless suffer from the abuse but start to consider their situation normatively appropriate. If our results generalize to environments outside the laboratory, they can explain a relative stability of corrupt institutions, as no party involved feels that anyone is doing anything inappropriate. It would also explain why removing those who abuse power is not enough, as others who fill their roles may act similarly, even if they were disadvantaged by that same behavior before. This happens because both parties follow what they believe is a social norm.

2 | Experimental Design

The general idea of our design is to model a social interaction where a minority can abuse institutionalized power. Our goal is to estimate how norms change after participants experience different levels of power abuse. Thus, to study the abusive behavior and the normative perception of power, we conducted a two-part experiment. The first part, aimed at creating a situation of power abuse, is very similar to the design used in Hoeft and Mill (2017) and in Hoeft and Mill (2024). In particular, a standard Public Goods game (the PGG) is implemented for 15 rounds with one subject randomly assigned to the additional role of a punisher throughout the game. The motivation behind using a PGG was to simulate a miniature community where every member can contribute to the common good, and one designated person, who is a regular member of the society, maintains order.[1] The second part utilizes the design of Krupka and Weber (2013) to elicit subjects' normative perceptions of different actions in the game (norm elicitation task). More specifically, subjects in power, and subjects not in power are asked to provide normative evaluations of several situations that could take place in the PGG.[2,3]

2.1 | Public Goods Game

All participants are randomly assigned a fixed role, either *punisher* or *nonpunisher*, and appointed to a group of four, in which

they remain for the 15 rounds of the PGG (partner matching). Each round of the PGG consists of three stages

Stage 1. Contribution to the Public Good. The first stage is a standard PGG (see the Supporting Information: Figure B.1 for a screenshot). Each of the four participants is endowed with 20 tokens and is asked to allocate this endowment between private and group accounts (1 token = 20 Euro cents). Tokens allocated to the private account are the subject's to keep. Tokens allocated to the group 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 $i \in \{A, ., D\}$ is defined as

$$\pi_i = 20 - c_i + 0.5 \cdot \sum_{j \in \{A,...,D\}} c_j \qquad (1)$$

Stage 2. Punishment. In the second stage (see the Supporting Information: Figure B.2 for a screenshot), the punishment decisions are made. While the three nonpunishing 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 in an anonymized way. To rule out reputation effects and to reduce the possibility of a punisher spitefully targeting individual nonpunishers, the information about nonpunishing participants is presented to the punisher in random order in an anonymized way in each round. Specifically, in each round anew, nonpunishers were randomly assigned the labels 1, 2, and 3 in this stage.

At this stage, the punisher is asked to indicate how many tokens he would like to deduct from the payoff of subject i (the amount deducted is denoted by σ_i, $i \neq D$).[4] 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 forfeited.[6] Thus, the punisher could reduce the payoff of the nonpunishers 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 with the contributions of others.

The payoff π_i of a nonpunisher $i \neq D$ is given by

$$\pi_i = 20 - c_i + 0.5 \cdot \sum_{j \in \{A,...,D\}} c_j - \sigma_i \qquad (2)$$

The payoff of the punisher is described by Equation (1). In the Supporting Information: Appendix A, we show that with selfish players the unique SPNE of this game is for the punisher to mete out maximum punishment of 10 tokens to each other player who does not contribute 20 tokens in any period, in which case all other players contribute optimally 20 tokens in each period and the punisher contributes zero in each period. Hence, a fully rational agent, focusing solely on monetary outcomes will punish others while refraining from contributing.[7]

Stage 3. Feedback. The third stage provides feedback to the participants (see the Supporting Information: Figure B.3 for a screenshot). 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 (as well as all other group members), which is common knowledge. This feedback ensures that group members can witness if there is abuse of power. Nonpunishers are not informed about the punishments meted out to others.

The design choices concerning feedback were driven by two objectives. On the one hand, we wanted to ensure that participants could spot and witness abuse of power, so that we can measure how this experience might affect future normative perceptions. Thus, to investigate how subjects react to experiences of power abuse, some feedback is essential. Therefore, all participants were informed about the contribution of punishers as well as all other group members. Further, participants were informed about the punishment they themselves received.[8]

On the other hand, we wanted to reduce the complexity and increase the chances of observing abusive behavior. Thus, we decided against the full information approach (i.e., nonpunishers observe the punishment of others on top of their own punishment). We did so for three reasons: (1) to not overload participants with information (as the experiment and the feedback are already rather complex), (2) to allow the punisher to try different punishment strategies (in particular in the beginning) without being constantly monitored, and (3) to have sufficient scope for punishers to behave abusively, which might have been reduced if they would have been more closely monitored by nonpunishers.

2.2 | Norm Elicitation Task

To elicit normative perceptions, 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). Thus, the norm elicitation task measures the injunctive norm. 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 modal response of other participants in this group, they are paid €8, otherwise they are paid €0.

There are two reasons why participants may deem actions of the punisher inappropriate: (1) undercontribution is inappropriate in general or (2) punishing while undercontributing is inappropriate. To disentangle these possibilities, we elicit the appropriateness of contributions by the punisher in two situations: full (FC-Q) and medium (MC-Q) contributions by the other group members. Additionally, we ask how appropriate it is to punish given different contributions by the punisher when the group contributes halfway (Pun-Q). The three situations, with the

corresponding five actions to be normatively assessed, are as follows:

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

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

Punishment Question **(Pun-Q)** Suppose the others (A, B, and C) contributed 10 tokens each to 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, and 20 tokens to the group account and reduces the payoff of A, B, or C.

In each of the three situations, subjects rate the social appropriateness of each action (contribution by D of 0, 5, 10, 15, and 20). For each action, the appropriateness is chosen on a 7-point Likert scale: *very socially inappropriate*, *socially inappropriate*, *somewhat socially inappropriate*, *neither appropriate nor inappropriate*, *somewhat socially appropriate*, *socially appropriate*, *very socially appropriate*.[9] To assess the social appropriateness of these situations, punishers indicate what level of appropriateness they think the mode of other punishers in the current session would choose (punishers' own reference group). Similarly, players A, B, and C (nonpunishers) indicate the level of appropriateness that they think the mode of other such players in the current session would choose (ABCs' own reference group). In addition, we collected the appropriateness perception of a third group of people. This group consisted of independent outsiders who did not participate in Part 1 of the experiment (the PGG), but were given the same instructions as punishers and nonpunishers. These subjects simply had to indicate the appropriateness levels that they thought the mode other independent outsiders in their session have chosen. Thus, these outsiders had not experienced the game, and their evaluations can be considered as the ex-ante normative perceptions of all players.[10]

2.3 | Payment

At the end of the experiment, subjects were paid for both tasks: the PGG and the appropriateness evaluation.[11] Subjects in the role of punishers and nonpunishers were paid for one randomly chosen round of the PGG. 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 mode of other subjects in the same role, she obtained €8, and zero otherwise. Overall, the average earning was €14.50 (including a show-up fee of €5).

2.4 | Procedure

Subjects 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.[12] After

taking part in the 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. The experiment lasted 1.5 hours (including seating, instructions, payoff, etc.). All measurements were computerized with the experimental software z-Tree (Fischbacher 2007).

The experiment was conducted in May 2018 at the Bonn DecisionLab and consisted of seven sessions that were conducted with subjects in the roles of punishers and nonpunishers (four sessions with 32 subjects and three sessions with 28 subjects). This sample size allows us to detect even medium-sized effects ($d = 0.48$) with a power of 80%. Three additional sessions totaling 77 subjects were conducted in the role of independent outsiders. Overall, 289 subjects (60% female) were recruited with the online registration software Hroot (Bock et al. 2014). The subjects' age ranged from 17 to 73 years (median = 22, mean = 24.14). Most were bachelor students (semester median = 3, mean = 4.18).[13]

3 | Hypotheses

Let us call subjects who played in role D in the PGG *punishers*, and subjects who played in roles A, B, and C *nonpunishers*. Our primary goal is to identify whether abuse of power becomes normalized for those who fall victim to it.

In order to do so, we need to specify which behavior constitutes power abuse in our design. Abuse of power is typically defined as a violation of norms governing its appropriate use (Grant and Keohane 2005). As there are no applicable legal norms in our setting, only social norms specify how the punishers should behave. As the punishers are randomly selected and incur no cost for punishing others, we define abuse of power as a combination of undercontribution and hypocritical punishment: contributing less than the nonpunishers but punishing them even if they contribute more than the punishers. In the Supporting Information: Appendix C.3, we show that there is broad consensus among the punishers, the nonpunishers and outsiders that both constitute a violation of social norms. As a second step, we need a strategy to identify the punishers who engage in abuse of power. Simply testing whether they engage in undercontribution and hypocritical punishment throughout our main task leads to problems of endogeneity: the punishers may react to those without power even though they have enough punishment points to enforce high contributions from every group member. We therefore classify the punishers based on their initial contribution in the PGG: specifically, on whether they make a below or above median first round contribution.[14] In the Supporting Information: Appendix C.2, we show that low initial contributions are predictive of future power abuse: punishers with low initial contributions contribute far less than the nonpunishers overall (undercontribution), enforce higher contributions from others (hypocritical punishment) and consequently earn more than the rest of their group. We will refer to nonpunishers randomly assigned to someone who abused power as *nonpunishers assigned an uncooperative punisher* and the rest *nonpunishers assigned a cooperative punisher*.

Our null hypothesis is that subjects have robust and common beliefs about social appropriateness of actions in the PGG, and that they are not distorted by any experiences in the game. Hence, under the null hypothesis, nonpunishers assigned an uncooperative and a cooperative punisher report the same social norms.

Hypothesis 0 *Regardless of their experience, the nonpunishers rate the social appropriateness of undercontribution and hypocritical punishment in the same way.*

The alternative hypothesis is that power, once abused, can corrupt the nonpunishers as well. The nonpunishers normalize such behavior even though they are disadvantaged by it (World Bank Group 2017). Nonpunishers assigned an uncooperative punisher believe that the prevailing social norm is more permissive than nonpunishers assigned a cooperative punisher.

Hypothesis 1 *Nonpunishers assigned an uncooperative punisher report higher social appropriateness ratings for undercontribution and hypocritical punishment than nonpunishers assigned a cooperative punisher.*

4 | Results

As already mentioned above (and similarly reported in Hoeft and Mill (2017) and in Hoeft and Mill (2024)), we find clear evidence of abusive behavior (see the Supporting Information: Appendix C.2 for a detailed analysis). Cooperative punishers act pro-socially and contribute on average about the same or even more than nonpunishers in their groups, and use punishment only restrictively. To the contrary, uncooperative punishers contribute substantially less than their nonpunishers, extensively use punishment, and obtain a substantially higher payoff than nonpunishers.

The behavior of uncooperative punishers is also consistently considered socially inappropriate by punishers, nonpunishers and outsides, as reported in detail in the Supporting Information: Appendix C.3.

However, the main question to be answered in this paper (which was impossible to answer with the data in Hoeft and Mill 2017 and in Hoeft and Mill 2024) is: How does the social norm perception of nonpunishers change after experiencing abusive behavior?

Before we get to our main result, notice that there are five elicited normative valences for each question expressed by the participants in our experiment. Specifically, in each question, participants report their perceived normative valences for five levels of hypothetical contributions by a punisher. In order to relate these evaluations to contributions and punishment levels in the PGG and to ease interpretation, we transform these into a single number. We do so by considering *average normative valences*. The average normative valence is just the average of the five normative valences expressed by a participant in a given question. The interpretation of the average normative valence differs slightly between each of the three questions. For the *full contributing question*, the average normative valence describes how socially acceptable undercontribution by the punisher is. For the *medium contributing question*, the average normative valence describes how socially acceptable undercontribution

by the punisher is if the nonpunishers also undercontributed (i.e., contributed 10). For the *punishment question*, the average normative valence describes how socially acceptable the punishment of undercontributing nonpunishers is if the punisher themselves undercontributed. Conceptually, the average normative valences of all these questions can be interpreted as how socially acceptable participants consider power abuse.

Notice that nonpunishers were assigned randomly to punishers who initially (i.e., in the first round) contribute below the median of other punishers (i.e., uncooperative punishers), and to punishers who initially contribute above or equal the median of other punishers (i.e., cooperative punishers). Therefore, any differences in norms that we detect between nonpunishers assigned an uncooperative or a cooperative punisher must be due to the experience that they had during the PGG. In fact, in the first round of the PGG the contributions of nonpunishers assigned an uncooperative or a cooperative punisher are statistically identical: they do not differ in their mean, median, minimum, or maximum contribution. Hence, all results for nonpunishers

can be *causally* attributed to the behavior of their punishers and the subsequent experience in the game. Specifically, from experiencing either a cooperative or an uncooperative punisher. This gives us an opportunity to see how the abusive behavior of uncooperative punishers and the cooperative behavior of cooperative punishers changes the nonpunishers' perception of the appropriateness of the punishers' actions.[15]

Figure 1 shows the nonpunishers' social norms as an average and as a function of punishers' contribution. The answers to the *full contributing question* (the middle panel of Figure 1) tell us what nonpunishers believe is the common attitude among the nonpunishers towards the *punishers'* free-riding. We see that nonpunishers assigned an uncooperative punisher consider it significantly *more* appropriate than nonpunishers assigned a cooperative punisher. This result is in support of Hypothesis 1: nonpunishers assigned an uncooperative punisher justify the low contributions of punishers by believing that this is socially appropriate. Nonpunishers assigned an uncooperative punisher also consider it significantly more appropriate than nonpunishers assigned a cooperative
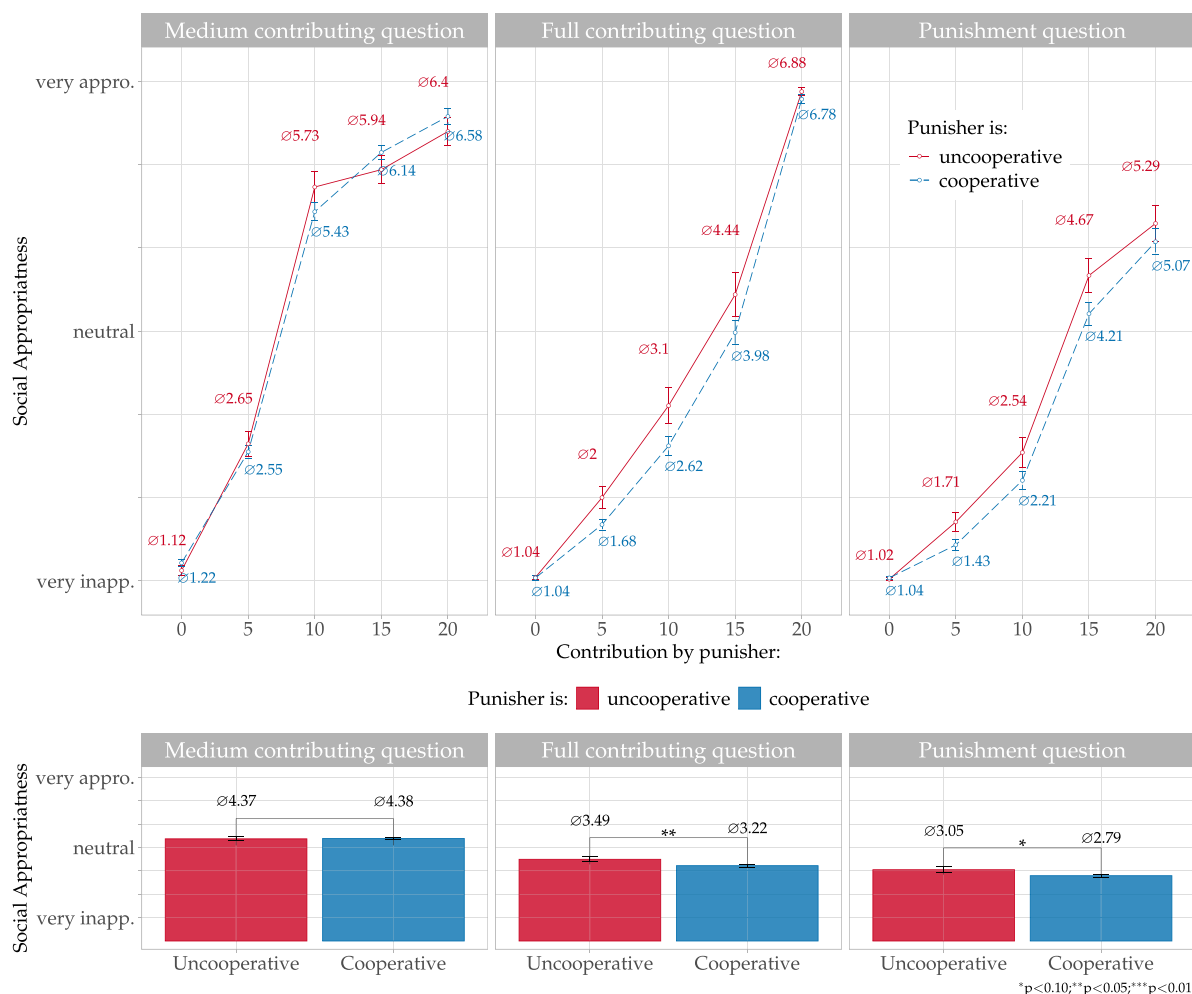

FIGURE 1 | Social norms reported by nonpunishers. *Note:* The figure depicts the social norm perception reported by nonpunishers. The top panels illustrate the social norm perception as a function of the punishers' contribution, while the bottom panels illustrate the average social norm perception. Left panels represent the normative valences for the *medium contributing question*, the mid-panels for the *full contributing question*, and the right panels for the *punishment question*. Blue bars and lines present the normative valences in the cooperative punisher-groups (i.e., the punisher contributed above the median in the first round of the PGG), while red bars and lines present the normative valences in uncooperative punisher-groups. Error bars denote standard errors. Hypothesis testing in the bottom panel was conducted using *t*-tests. Stars indicate significance at the 10% (*), 5% (**), and 1% (***) level. [Colour figure can be viewed at wileyonlinelibrary.com]

punisher when punishers subtract money from them (punishment question, Pun-Q; right panel of Figure 1).

Table 1 demonstrates the same point with a regression that treat punishers' initial contribution as a continuous variable. Importantly, unlike punishers, *the nonpunishers are not those who punish, but those who receive the punishment.* Therefore, nonpunishers assigned an uncooperative punisher—instead of seeing the hypocritical punishment, which comes from a person who contributes less than them, as "unfair" and thus inappropriate—start to believe that it is actually justified to abuse the power (Hypothesis 1). Specifically, nonpunishers assigned an uncooperative punisher consider undercontribution and punishment while undercontributing more appropriate than nonpunishers assigned a cooperative punisher.[16]

It is important to note that we adopt a conservative approach to estimating the effect. Specifically, we use the punisher's

TABLE 1 | Estimation of the social normative perception.

Panel A: Social norms as function of punisher's contribution

	FC-Q	MC-Q	Pun-Q	FC-Q	MC-Q	Pun-Q	FC-Q	MC-Q	Pun-Q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	3.44***	4.40***	3.09***	3.74***	4.36***	3.20***	3.82***	4.44***	3.37***
	(0.13)	(0.08)	(0.12)	(0.18)	(0.12)	(0.18)	(0.19)	(0.13)	(0.19)
$\text{Cont.Pun}_{t\in\{1\}}$	−0.01	−0.002	−0.02**						
	(0.01)	(0.01)	(0.01)						
$\text{Cont.Pun}_{t\in\{1,\dots,5\}}$				−0.03***	0.001	−0.02**			
				(0.01)	(0.01)	(0.01)			
$\text{Cont.Pun}_{t\in\{1,\dots,15\}}$							−0.03***	−0.004	−0.03***
							(0.01)	(0.01)	(0.01)
Group specific effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Log likelihood	−166.52	−123.65	−191.74	−163.6	−123.42	−191.63	−162.97	−123.28	−189.71
Observations	159	159	159	159	159	159	159	159	159

Panel B: Social norms as function of how often the punisher undercontributed and punished

	FC-Q	MC-Q	Pun-Q	FC-Q	MC-Q	Pun-Q	FC-Q	MC-Q	Pun-Q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	3.22***	4.39***	2.77***	3.26***	4.39***	2.81***	3.14***	4.35***	2.76***
	(0.08)	(0.05)	(0.08)	(0.07)	(0.05)	(0.07)	(0.08)	(0.05)	(0.08)
$\#\{c_p^t < c_{-p}^{t-1}\}$	0.02	−0.004	0.03*						
	(0.01)	(0.01)	(0.01)						
$\#\{c_p^t < c_{-p}^{t-1}\} \geq 10$				0.27	−0.08	0.38**			
				(0.18)	(0.12)	(0.17)			
$\#\{\sigma_i > 0\}$							0.05***	0.01	0.03**
							(0.01)	(0.01)	(0.02)
Group specific effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Log likelihood	−165.71	−123.17	−191.63	−163.13	−120.45	−188.51	−160.72	−122.74	−191.19
Observations	159	159	159	159	159	159	159	159	159

Note: The table depicts the average social norm perception as a function of punisher's contributions, how often the punisher undercontributed, and how often the punisher punished in the public good game. *FC-Q, MC-Q,* and *Pun-Q* denote the average normative valence in the *full contributing question, medium contributing question* and the *punishment question*, respectively. $\text{Cont.Pun}_{t\in\{1\}}$ denotes the punishers' contribution (continuous measure) in the first round. $\text{Cont.Pun}_{t\in\{1,\dots,15\}}$ denotes the average punishers' contribution (continuous measure) in the first fifteen rounds, and $\text{Cont.Pun}_{t\in\{1,\dots,5\}}$ is defined accordingly. $\#\{c_p^t < c_{-p}^{t-1}\}$ denotes the number of times the punisher contribute in period t less than the average nonpunisher contributed in $t-1$. $\#\{c_p^t < c_{-p}^{t-1}\} \geq 10$ denotes a dummy with value one if the punisher at least 10 times contribute in period t less than the average nonpunisher contributed in $t-1$. $\#\{\sigma_i > 0\}$ denotes how often a nonpunisher has received punishment. Heterogeneity on the group level is accounted for by group-specific random-intercept effects.
*p<0.10;**p<0.05;***p<0.01;****p<0.001.

first-round contribution as an instrument to causally estimate the impact of abuse on social norms. While this method has the advantage of providing clean and causal insights, it also has limitations. Specifically, the punisher's behavior in the first round may not accurately reflect their behavior in subsequent rounds, potentially affecting the overall experience of nonpunishers. For instance, a punisher who contributes nothing in the first round but contributes fully in all subsequent rounds may provide non-punishers with a predominantly positive experience, which would not be captured by their initial behavior. Conversely, a punisher who contributes generously in the first round but ceases contributions in later rounds could create the opposite effect. These dynamics suggest that a more comprehensive measure, accounting for the intensity and duration of undercontribution as well as punishment throughout the game, might better capture the true impact of abuse.

We can see from Table 1 (columns 4–9) how more exposure to hypocritical behavior of punishers affects the nonpunishers' average normative valences. First, we see that focusing on the average contribution of punishers in the first five, and also all rounds, result in similar conclusions. More interestingly, we see that the level of the effect *grows* in magnitude and in significance. Thus, we find the strongest effect of nonpunishers assigned an uncooperative punisher considering punishers' free-riding significantly more appropriate than nonpunishers assigned a cooperative punisher if we focus on the average punishers' contributions over the whole game.

In addition, panel B of Table 1 reports how the frequency of nonpunishers being punished affects their normative valences. Nonpunishers who have been punished more often consider undercontribution and punishment again *more* socially appropriate. Similarly, we see from Table 1 that the number of times a punisher contributes less than their nonpunishers contribution in the previous round also changes the normative valences. Specifically, nonpunishers who experienced a punisher who undercontributed more often consider it, again, *more* appropriate to punish and undercontribute. At the same time, we need to point out that focusing on the average punishers' contribution over the whole game, or the number of times nonpunishers have been punished, as well as the number of times punishers undercontribute might have issues with endogeneity as punishers might, in part, react to nonpunishers' behavior. Thus, we obtain the cleanest (and clearly most unconfounded) results, if we focus on the punishers' initial contributions.

One concern a critical reader might have is that even if the non-punishers assigned an uncooperative punisher report higher social appropriateness ratings for abuse of power, they may simply do so because they believe their reference group was corrupted. Our method of elicitation builds on a coordination task that is locally and temporally contingent on the reference group and its experiences. Even if those assigned an uncooperative punisher remain convinced of certain behaviors being socially inappropriate in general, they may believe that for their reference group the social norm has lost its salience due to competing descriptive norms. If this were true, we would expect no differences when eliciting the social appropriateness of nonpunishers when incentivized to coordinate with participants who have not played the game but just learned the rules of the

game (i.e., outsiders). However, we still observe that nonpunishers assigned an uncooperative punisher consider undercontribution ($\beta = -0.01$, $t(159) = -1.73$, $p = 0.085$) and punishment ($\beta = -0.02$, $t(159) = -2.00$, $p = 0.047$) again *more* socially appropriate than nonpunishers assigned a cooperative punisher, even if they are incentivized to coordinate with outsiders.[17] This suggests that our results are not just an artifact of the coordination device, but genuinely represents the social norm perception of nonpunishers.[18]

Bringing these insights together reveals the following pattern. Nonpunishers who experience a punisher who initially undercontributed perceive undercontribution as more socially acceptable compared with nonpunishers who experienced a punisher who initially overcontributed. When the punisher undercontributes throughout the entire game, nonpunishers perceive undercontribution as even more acceptable than those who experienced a punisher who consistently overcontributed. Furthermore, nonpunishers subjected to frequent punishment by the punisher are more likely to view undercontribution as socially acceptable than those who experienced a punisher who was more restrained in their use of punishment. In summary, experiencing an abusive punisher systematically shifts non-punishers' beliefs about what others perceive as right or wrong, making them more accepting of abusive behavior.

5 | Limitations and Discussion

It is important to acknowledge that our experimental design deliberately permits the exercise of power without imposing direct costs, a feature that may lead some to question whether what we label as "abuse of power" is indeed abusive, or merely expected sanctioned behavior. In our setting, the punisher operates within a framework that explicitly allows for, and even incentivizes, behavior that might be considered exploitative. This design choice purposefully serves to induce abusive behavior and to isolate the effects of repeated exposure to such behavior on normative perceptions. However, by situating the behavior within an institutionally sanctioned environment, one might argue that the observed actions are simply the expected outcome of the rules rather than an inherent violation of ethical or moral norms. While, this limitation is important, we believe that this issue is likely true for most experiments and potentially inherent to experimental research with volunteer participants. Any laboratory experiment needs to offer fixed action sets to participants. To explore abusive behavior, we need to create an institution which permits such behavior. Even if we imposed sanctions (e.g., through peer or third-party punishment), we would still need a framework that explicitly permits the actions under study. Thus, it may be conceptually unfeasible to examine abusive behavior in a controlled setting without facilitating it.

A related concern could be that participants might consider the setting itself objectionable while not considering it objectionable if punishers act according to what is possible in the setting. Our paper only speaks to whether the act of power abuse is considered objectionable in one specific setting, where power abuse is institutionally sanctioned. It, however, does not speak to whether the setting itself is objectionable. While we believe

it is important to study whether and when abusive behavior is considered socially inappropriate, it is also important to study how the institution facilitating abusive behavior is considered. Further research should study whether also the normative perception of the institutional setting itself evolves as a function of experienced power abuse.

It should also be highlighted that it is not particularly surprising that abusive punishers perceive it as appropriate to exploit their power. For one, the experiment is specifically designed to induce abusive behavior, intentionally permitting and potentially legitimizing such behavior. In addition, punishers self-select into being abusive or not. As a result, those who are predisposed to find such actions acceptable are more likely to behave abusively. However, the primary focus of this paper is on the attitudes of nonpunishers, who have randomly been assigned to either an abusive or less abusive punisher. Although the institutional framework legitimizes the execution of power within the game, our findings reveal that the nonpunishers' normative judgments significantly shift depending on their exposure to abusive behavior.

A related issue could be that our results resemble a framing effect, as the behavior in question is not necessarily inappropriate in itself, but rather, it is the institutional rule that permits such behavior that could be seen as problematic. It, indeed, might be the case that the institutional rules that permit such behavior could shape perceptions independently of the behavior itself. However, if the appropriateness questions were capturing only the institutional framework, we would expect limited variation in responses across different behaviors within the same institutional context (i.e., for each of the three questions and the five possible behaviors of punishers). In contrast, our data show strong variation both between subjects (depending on whether they interacted with an abusive or less abusive punisher) and within subjects across the different norm elicitation questions. This pattern suggests that the evaluative shifts are primarily driven by the experienced behavior rather than a static perception of the institutional rules. In other words, while the framing provided by the institution might influence initial perceptions, the change in normative evaluations appears to be a direct function of the abusive behavior experienced, supporting the interpretation that it is the behavior—and not merely the rule that permits it—that alters social norm perceptions.

6 | Conclusion

The corrupting effect of power on those who wield it has been studied extensively (e.g., Fehr et al. 2013; Falk and Kosfeld 2006; Hoeft and Mill 2017). We show how abuse of power can also corrupt the powerless even though they fall victim to this abuse. Our research design implements a repeated Public Goods game in which only one member of the group (the punisher) has the power to costlessly punish others (the nonpunishers). After the game, we measure the normative beliefs of the punisher, the nonpunishers and outsiders with respect to the actions of the punisher. We find broad consensus on the (in)appropriate use of power. Undercontributing while punishing higher contributions (hypocritical punishment) is seen as a violation of social norms by all parties. More importantly, our results show that

those who fall prey to abusive behavior, are more inclined to consider undercontribution and hypocritical punishment to be socially appropriate compared with those who were spared abusive behavior.

Our findings are likely to underreport the extent of the problem in applied settings. In our experiment, the powerful (punishers) are chosen randomly, whereas in the real world people with power are often chosen through some measure of merit or some form of voting. This may additionally legitimize the actions of punishers and make it more plausible for the nonpunishers to start believing that abusive behavior is backed by a social norm. People in power can be perceived as having better knowledge with regard to what should be done, which makes any actions that they take look more appropriate than these actions may actually be.

Overall, our results unveil a mechanism that might be responsible for many failed attempts to fight corruption on domestic and international levels, and point toward a reason why inefficient institutions endure. On the one hand, people in power may abuse it because they do not find anything wrong with such behavior. On the other hand, people who are being abused start believing that this is a social norm and may not voice their concerns. Thus, the phenomenon we have uncovered can contribute to the stability of corrupt institutions.

To counter these dynamics, several policy measures may offer promising avenues. Fostering public debates and forums can enable the emergence of counter-narratives from those affected by abuse, challenging the normative acceptance of such practices. Similarly, enhanced transparency—such as routine disclosure of institutional actions—can expose and simplify the redress of abusive behavior. Additionally, introducing rotating leadership roles and enforced term limits may prevent the consolidation of power, mitigating its corrupting effects. Such mechanisms may help prevent abusive practices from becoming normalized and even begin to reverse entrenched norm shifts. It is important to note, however, that we do not provide any evidence of these policy suggestions being effective, so they should be interpreted with caution and only seen as a starting point for further policy development rather than definitive solutions.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Endnotes

[1] Note that this is a rather conservative implementation to explore power abuse as in other experiments, and in reality most people in power achieve it by some kind of merit or voting based procedures.

[2] Subjects only learned the nature of the task in the second part after the first part was concluded.

[3] We deliberately put the norm elicitation task only at the end of the experiment, instead of having it before and after the PGG (which seems reasonable if one wants to detect changes in normative beliefs). There are several reasons for this choice: First, it has been shown that repeating the same task, and specifically answering *the same questions*, activates a drive for consistency and, thus, may dilute the studied effects (Johansson-Stenman and Svedsäter 2008). Second, a repeated norm elicitation task asking the same questions might be prone to demand effects, as subjects may reason that the study is about changes in normative perceptions. Third, asking for the normative evaluation of situations before the PGG might frame punishers to behave less abusively, which might substantially reduce the participants' experience of power abuse. All these issues related to internal validity contributed to our choice of measuring normative beliefs only after the PGG.

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

[5] Note that the highest individual benefit from free-riding when the other two nonpunishers contribute 20 tokens, is 10 tokens. If a punisher was 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 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 nonpunishers. 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 nonpunishers incomparable.

[7] Note that a different way to create "power" would be to have a sequential stage game, similar to Gächter and Renner (2018). Specifically, the punisher could also be a first-mover. However, we decided against it, as it would mean an additional difference between the punisher and nonpunishers. Our goal was to make power abuse as straightforward as possible. Thus, we wanted the punisher to differ from nonpunishers solely in their power to punish, while they should be comparable in the contribution and feedback stage. Hence, any differences in the contribution stage would have made power abuse less clear, which is why we decided to induce power solely through a punishment opportunity.

[8] Specifically, the feedback about contributions was necessary so that the nonpunishers could directly observe the punisher using his power. If we were to give no feedback at all, there would be no experience of abuse, as the hypocrisy of punishers would not be observed. If we were to only report averages, as opposed to the individual contributions of all participants, this would obfuscate the actions of the punisher, and we would have to control for nonpunishers' beliefs as they may think the punisher actually contributes a lot, and only nonpunishers refuse to contribute.

[9] We chose seven instead of five statements as originally used by Krupka and Weber (2013) (see the Supporting Information: Tables B.1, B.2, and B.3 for further details). The main reason to do so was to ensure sufficient variation in the data and to eliminate a very clear focal point (which might result in a possible demand effect). Specifically, having only five appropriateness statements for each of the five actions would make it salient and likely that participants would answer diagonally, that is, choosing different appropriateness levels for each of the five actions. Such a design decision could potentially reduce variation and bias results by providing a very salient artificial focal point. Using seven instead of five statements reduces this issue.

[10] In the experiment, punishers/nonpunishers and outsiders were also asked to evaluate the levels of appropriateness chosen by the mode of the nonpunishers/punishers and outsiders in the current session. We discuss these data in a separate paper.

[11] Following the arguments of Charness et al. (2016) and Azrieli et al. (2018), we decided not to pay for all decisions in the experiment to reduce hedging. At the same time, we did not want to dilute the incentives, in particular in the norm elicitation task. Thus, to find the right balance, we incentivized the three tasks separately (for an overview of other papers using such an approach, see Charness et al. 2016). We believe that this did not create any problems with hedging as there was no feedback between the tasks. Moreover, before the experiment, subjects were only informed that there would be three tasks (without knowing whether and how the upcoming tasks will be incentivized).

[12] The instructions as well as an English version of the handout can be found in the supporting information.

[13] See the Supporting Information: Section C.1 for more detailed sample characteristics.

[14] We use the median split explicitly for illustrative purposes. We also replicate all our results using a continuous scale of punishers' initial contributions. Thus, the results below do not hinge on the median split of the groups.

[15] Note that we also observe the social norm perception of punishers. However, their norm perception is endogenous to their behavior in the game. In fact, we find that uncooperative punishers consider it more appropriate than cooperative punishers to free-ride after others have contributed the full amount. A similar difference can be observed for the *punishment question*, where uncooperative punishers consider it more appropriate than cooperative punishers to punish a nonpunisher's contribution of 10 tokens while contributing small amounts themselves. However, as punishers self-select into their first-round behavior, we cannot disentangle whether abusing others changes the punishers' social norm perception, or whether a different perception of the social norms affects the punishers' behavior in the PGG.

[16] To make use of the detailed data on normative valences for all five hypothetical contributions of the punisher we can also focus on the slope of the norm function formed by normative valences. As the norm functions are nonlinear we can use a generalized additive model (GAM) to estimate the effect of cooperative vs. uncooperative punishers on normative valences. All our results can be replicated using slopes.

[17] This effect, again, becomes stronger if we consider the undercontribution throughout the whole experiment instead of the punishers' initial contribution.

[18] In the Supporting Information: Appendix C.4, we report on the results of a follow-up experiment where we remove the incentives to coordinate completely. While we observe no differences in norms, the sample was collected at a different location during COVID-19, warranting caution in interpreting these findings.

References

Acemoglu, D., and J. A. Robinson. 2008. "Persistence of Power, Elites, and Institutions." *American Economic Review* 98, no. 1: 267–293.

Acemoglu, D., and J. A. Robinson. 2019. *The Narrow Corridor: States, Societies, and the Fate of Liberty: Winners of the 2024 Nobel Prize in Economics.* Penguin UK.

Alger, I., S. Gavrilets, and P. Durkee. 2024. "Proximate and Ultimate Drivers of Norms and Norm Change." *Current Opinion in Psychology* 60: 101916.

Álvarez-Benjumea, A. 2022. "Uncovering Hidden Opinions: Social Norms and the Expression of Xenophobic Attitudes." *European Sociological Review* 39, no. 3: 449463.

Asch, S. E. 1956. "Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority." *Psychological Monographs: General and Applied* 70, no. 9: 170.

Ashforth, B. E., and V. Anand. 2003. "The Normalization of Corruption in Organizations." *Research in Organizational Behavior* 25: 1–52.

Azrieli, Y., C. P. Chambers, and P. J. Healy. 2018. "Incentives in Experiments: A Theoretical Analysis." *Journal of Political Economy* 126, no. 4: 1472–1503.

Banerjee, R. 2016. "Corruption, Norm Violation and Decay in Social Capital." *Journal of Public Economics* 137: 14–27.

Basić, Z., and E. Verrina. 2024. "Personal Norms — and Not Only Social Norms — Shape Economic Behavior." *Journal of Public Economics* 239: 105255.

Bendahan, S., C. Zehnder, F. P. Pralong, and J. Antonakis. 2015. "Leader Corruption Depends on Power and Testosterone." *Leadership Quarterly* 26, no. 2: 101–122.

Bicchieri, C., E. Dimant, S. Gächter, and D. Nosenzo. 2022. "Social Proximity and the Erosion of Norm Compliance." *Games and Economic Behavior* 132: 59–72.

Bicchieri, C., and L. Garzino Demo. 2025. "Norm Strength and Norm Stability." *Current Opinion in Psychology* 62: 101957.

Bicchieri, C., and E. Xiao. 2009. "Do the Right Thing: But Only If Others Do So." *Journal of Behavioral Decision Making* 22, no. 2: 191–208. https://doi.org/10.1002/bdm.621.

Bock, O., I. Baetge, and A. Nicklisch. 2014. "hroot: Hamburg Registration and Organization Online Tool." *European Economic Review* 71, no. C: 117–120.

Charness, G., U. Gneezy, and B. Halladay. 2016. "Experimental Methods: Pay One or Pay All." *Journal of Economic Behavior & Organization* 131: 141–150.

Cialdini, R. B. 2013. *Influence: Pearson New.* International ed. *Science and Practice.*

Cialdini, R. B., and N. J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55, no. 1: 591621.

Cialdini, R. B., R. R. Reno, and C. A. Kallgren. 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology* 58, no. 6: 1015–1026.

Dal Bó, E., P. Dal Bó, and J. Snyder. 2009. "Political Dynasties." *Review of Economic Studies* 76, no. 1: 115–142.

Dubois, D., D. D. Rucker, and A. D. Galinsky. 2015. "Social Class, Power, and Selfishness: When and Why Upper and Lower Class Individuals Behave Unethically." *Journal of Personality and Social Psychology* 108, no. 3: 436–449.

Falk, A., and M. Kosfeld. 2006. "The Hidden Costs of Control." *American Economic Review* 96, no. 5: 1611–1630.

Fehr, E., H. Herz, and T. Wilkening. 2013. "The Lure of Authority: Motivation and Incentive Effects of Power." *American Economic Review* 103, no. 4: 1325–1359.

Fischbacher, U. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10, no. 2: 171–178.

Fischbacher, U., S. Gächter, and E. Fehr. 2001. "Are People Conditionally Cooperative? Evidence From a Public Goods Experiment." *Economics Letters* 71, no. 3: 397–404.

Fisman, R., and E. Miguel. 2007. "Corruption, Norms, and Legal Enforcement: Evidence From Diplomatic Parking Tickets." *Journal of Political Economy* 115, no. 6: 1020–1048.

Furnham, A. 2003. "Belief in a Just World: Research Progress Over the Past Decade." *Personality and Individual Differences* 34, no. 5: 795–817.

Gächter, S., and E. Renner. 2018. "Leaders as Role Models and 'Belief Managers' in Social Dilemmas." *Journal of Economic Behavior & Organization* 154: 321–334.

Gächter, S., and J. F. Schulz. 2016. "Intrinsic Honesty and the Prevalence of Rule Violations Across Societies." *Nature* 531, no. 7595: 496–499.

Galinsky, A. D., D. D. Rucker, and J. C. Magee. 2015. "Power: Past Findings, Present Considerations, and Future Directions." In *APA Handbook of Personality and Social Psychology, Volume 3: Interpersonal Relations*, 421–460. American Psychological Association.

Gelfand, M. J., S. Gavrilets, and N. Nunn. 2024. "Norm Dynamics: Interdisciplinary Perspectives on Social Norm Emergence, Persistence, and Change." *Annual Review of Psychology* 75, no. 1: 341378.

Giurge, L. M., M. van Dijke, M. X. Zheng, and D. De Cremer. 2021. "Does Power Corrupt the Mind? The Influence of Power on Moral Reasoning and Self-Interested Behavior." *Leadership Quarterly* 32, no. 4: 101288.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman. 1996. "Crime and Social Interactions." *Quarterly Journal of Economics* 111, no. 2: 507–548.

Grant, R. W., and R. O. Keohane. 2005. "Accountability and Abuses of Power in World Politics." *American Political Science Review* 99, no. 1: 29–43.

Greiner, B. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments With ORSEE." *Journal of the Economic Science Association* 1: 1–12.

Guinote, A., I. Cotzia, S. Sandhu, and P. Siwa. 2015. "Social Status Modulates Prosocial Behavior and Egalitarianism in Preschool Children and Adults." *Proceedings of the National Academy of Sciences* 112, no. 3: 731–736.

Herz, H., and D. Taubinsky. 2017. "What Makes a Price Fair? An Experimental Study of Transaction Experience and Endogenous Fairness Views." *Journal of the European Economic Association* 16: 1–37.

Hoeft, L., M. Kurschilgen, and W. Mill. 2025. "Norms as Obligations." *International Review of Law and Economics* 81: 106235.

Hoeft, L., and W. Mill. 2017. "Selfish Punishers." *Economic Letters* 157: 41–44.

Hoeft, L., and W. Mill. 2024. "Abuse of Power: An Experimental Investigation of the Effects of Power and Transparency on Centralized Punishment." *Journal of Economic Behavior & Organization* 220: 305–324.

Johansson-Stenman, O., and H. Svedsäter. 2008. "Measuring Hypothetical Bias in Choice Experiments: The Importance of Cognitive Consistency." *B.E. Journal of Economic Analysis & Policy* 8, no. 1: 1–8.

Kipnis, D. 1972. "Does Power Corrupt?" *Journal of Personality and Social Psychology* 24, no. 1: 33–41.

Klitzman, R. 2007. *When Doctors Become Patients.* Oxford University Press Inc.

Krupka, E. L., and R. A. Weber. 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of European Economic Association* 11, no. 3: 495–524.

Lammers, J., A. D. Galinsky, D. Dubois, and D. D. Rucker. 2015. "Power and Morality." *Current Opinion in Psychology* 6: 15–19.

Lammers, J., D. A. Stapel, and A. D. Galinsky. 2010. "Power Increases Hypocrisy: Moralizing in Reasoning." *Immorality in Behavior. Psychological Science* 21, no. 5: 737–744.

Lindström, B., S. Jangard, I. Selbing, and A. Olsson. 2018. "The Role of a "Common Is Moral" Heuristic in the Stability and Change of Moral Norms." *Journal of Experimental Psychology: General* 147, no. 2: 228–242.

Mallucci, P., D. Y. Wu, and T. H. Cui. 2019. "Social Motives in Bilateral Bargaining Games: How Power Changes Perceptions of Fairness." *Journal of Economic Behavior & Organization* 166: 138–152.

Merguei, N., M. Strobel, and A. Vostroknutov. 2022. "Moral Opportunism as a Consequence of Decision Making Under Uncertainty." *Journal of Economic Behavior & Organization* 197: 624–642.

Mooijman, M., W. W. van Dijk, N. Ellemers, and E. van Dijk. 2015. "Why Leaders Punish: A Power Perspective." *Journal of Personality and Social Psychology* 109, no. 1: 75–89.

Olken, B. A., and R. Pande. 2012. "Corruption in Developing Countries." *Annual Review of Economics* 4, no. 1: 479–509.

Robinson, J. A., Vostroknutov, A., and Vostroknutova, E. (2023). *Endogenous Institutions and Economic Policy. Policy Research Working Paper Series 10600*, The World Bank.

Rustichini, A., and M. C. Villeval. 2014. "Moral Hypocrisy, Power and Social Preferences." *Journal of Economic Behavior & Organization* 107: 10–24.

Schier, U. K., A. Ockenfels, and W. Hofmann. 2016. "Moral Values and Increasing Stakes in a Dictator Game." *Journal of Economic Psychology* 56: 107–115.

Schilke, O., M. Reimann, and K. S. Cook. 2015. "Power Decreases Trust in Social Exchange." *Proceedings of the National Academy of Sciences* 112, no. 42: 12950–12955.

Smerdon, D., T. Offerman, and U. Gneezy. 2020. "Everybodys Doing It: On the Persistence of Bad Social Norms." *Experimental Economics* 23, no. 2: 392420.

Tabellini, G. 2008. "Institutions and Culture." *Journal of the European Economic Association* 6, no. 2-3: 255–294.

Tabellini, G. 2010. "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association* 8, no. 4: 677–716.

Tyler, T. R. 2000. "Social Justice: Outcome and Procedure." *International Journal of Psychology* 35, no. 2: 117125.

Tyler, T. R. 2006. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57, no. 1: 375400.

Tyler, T. R., and S. L. Blader. 2003. "The Group Engagement Model: Procedural Justice, Social Identity, and Cooperative Behavior." *Personality and Social Psychology Review* 7, no. 4: 349361.

Tyler, T. R., and E. A. Lind. 1992. *A Relational Model of Authority in Groups*, 115191. Elsevier.

van Dijk, E., D. De Cremer, and M. J. J. Handgraaf. 2004. "Social Value Orientations and the Strategic Use of Fairness in Ultimatum Bargaining." *Journal of Experimental Social Psychology* 40, no. 6: 697–707.

van Kleef, G. A., F. Wanders, E. Stamkou, and A. C. Homan. 2015a. "The Social Dynamics of Breaking the Rules: Antecedents and Consequences of Norm-Violating Behavior." *Current Opinion in Psychology* 6: 25–31.

van Kleef, G. A., F. Wanders, E. Stamkou, and A. C. Homan. 2015b. "The SOCIAL DYNamics of breaking the Rules: Antecedents and Consequences of Norm-Violating Behavior." *Current Opinion in Psychology* 6: 2531.

Vredenburgh, D., and Y. Brender. 1998. "The Hierarchical Abuse of Power in Work Organizations." *Journal of Business Ethics* 17, no. 12: 1337–1347.

Werner, B., W. Mill, and W. Hariskos. 2024. "Moneyball: Studying Effects of Ostracism on Both Targets and Voluntary Sources." *Social Psychology* 55, no. 5: 229–242.

Wiltermuth, S. S., and F. J. Flynn. 2013. "Power, Moral Clarity, and Punishment in the Workplace." *Academy of Management Journal* 56, no. 4: 1002–1023.

Wong, K. C. 1998. "A Reflection on Police Abuse of Power in the People's Republic of China." *Police Quarterly* 1, no. 2: 87–112.

World Bank Group. 2017. *World Development Report 2017: Governance and the Law*. World Bank.

Xu, A. J., R. Loi, and L. W. Lam. 2015. "The Bad Boss Takes It All: How Abusive Supervision and Leader–Member Exchange Interact to Influence Employee Silence." *Leadership Quarterly* 26, no. 5: 763–774.

Zenou, Y. 2003. "The Spatial Aspects of Crime." *Journal of the European Economic Association* 1, no. 2-3: 459–467.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.