Hermstrüwer, Yoan; Khesali, Mahdi

Working Paper

The democracy premium in expressive law: An experiment

Discussion Papers of the Max Planck Institute for Research on Collective Goods, No. 2025/6

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

Suggested Citation: Hermstrüwer, Yoan; Khesali, Mahdi (2025) : The democracy premium in expressive law: An experiment, Discussion Papers of the Max Planck Institute for Research on Collective Goods, No. 2025/6, Max Planck Institute for Research on Collective Goods, Bonn, https://hdl.handle.net/21.11116/0000-0011-2D24-3

This Version is available at:
https://hdl.handle.net/10419/318544

MAX PLANCK INSTITUTE
FOR RESEARCH ON
COLLECTIVE GOODS

YOAN HERMSTRÜWER
MAHDI KHESALI

Discussion Paper
2025/6

THE DEMOCRACY PREMIUM IN EXPRESSIVE LAW: AN EXPERIMENT

The Democracy Premium in Expressive Law*

An Experiment

Yoan Hermstrüwer† Mahdi Khesali‡

April 30, 2025

Abstract

Why do people obey the law when it is not formally enforced? In this study, we explore the expressive power of democracy as a behavioral channel of compliance with the law. Using a modified version of the stealing game, we examine the effect of two distinct democratic interventions on stealing under normative ambiguity: a voting procedure in which the outcome is revealed, and a voting procedure in which the outcome of the vote remains unknown. We find that revealing the outcome of a vote significantly reduces stealing relative to a baseline treatment without a vote and the treatment in which the outcome of the vote remains unknown. We also observe suggestive evidence that participants who support the social norm proscribing theft are more likely to steal nonetheless when the outcome remains unknown. Our findings have important implications for the design of expressive law and of democratic voting procedures.

JEL: C91; D72; D91; K14; K42

*We thank Christoph Engel, Pascal Langenbach, Adi Leibovitch, Doron Teichman, Eyal Zamir, and participants at the Annual Conference of the European Association of Law and Economics (EALE 2024), at the Annual Conference of the European Society for Empirical Legal Studies (ESELS 2024), and at the Regulation Research Conference 2023 in Regensburg for helpful comments and discussions. This project was funded by the Max Planck Institute for Research on Collective Goods. The authors declare no conflict of interest. This study was approved under the general approval agreement of the Decision Lab at the Max Planck Institute for Research on Collective Goods. We thank Michael Seebauer for his invaluable assistance in running the experiment. No material from other sources was used.

†University of Zurich. E-mail: yoan.hermstruewer@ius.uzh.ch

‡University of Hamburg and Max Planck Institute for Research on Collective Goods. E-mail: mahdi.khesali@ile-hamburg.de

1 Introduction

Why do individuals, corporations, and entire states so frequently follow legal norms – even when there is little or no formal or informal enforcement to ensure compliance? Louis Henkin, a pioneering figure in modern human rights law, noted that "almost all nations observe almost all principles of international law and almost all of their obligations almost all of the time" (Henkin, 1979). This pattern is also evident domestically: people buckle their seatbelts when getting into a car; devoted *bon vivants* refrain from smoking in public; and petty crime remains relatively rare in many societies, even when the likelihood of being caught is slim. These observations raise an intriguing conundrum about the cognitive and motivational forces driving compliance with the law when the threat of punishment is minimal or nonexistent.

One influential explanation in legal scholarship attributes law-abiding behavior to the expressive function of law (Cooter, 1998; McAdams, 2015; Nadler, 2024; Sunstein, 1996). According to this view, legal rules can shape behavior by conveying a normative signal about what is socially or legally appropriate, clarifying the meaning of certain actions in the broader legal and social context. A key aspect of this communicative process is its capacity to reduce normative ambiguity, which arises when multiple norms prescribing different courses of action coexist in a given context (Engel et al., 2021). Indeed, proponents of expressive law argue that normative ambiguity is a necessary condition for the expressive effect of the law to materialize (Teichman & Leibovitch, 2024). These expressive law theories emphasize the interplay between legal and social norms. Social norms can be either descriptive – capturing what people typically do (their *consuetudo*) – or injunctive – representing what the majority deems morally or socially authoritative (Bicchieri, 2005; Cialdini et al., 1990).

A critical question is how best to elicit the appropriate course of action so as to effectively steer behavior. A recent line of research on self-nudging and behavioral self-management (cf. Tontrup & Sprigman, 2022) suggests that simply prompting people to reflect on their actions may steer them towards normative compliance without requiring explicit knowledge of a broader social consensus. This approach, relying on informal institutions and personal introspection, contrasts with more structured modes guiding the production of norms in democratic societies. In democracies, norms are often shaped by formal collective decision-making processes such as elections or referenda. Switzerland, for instance, is notable for its direct democratic practices, in which citizens regularly vote on political issues and legislation.

Democracy has been consistently associated with higher levels of cooperation and compliance, a phenomenon sometimes referred to as the democracy premium (Dal Bó et al., 2010). However, it is less clear which specific features of democratic procedures drive this effect. On the one hand, public outcomes generate social proof by visibly demonstrating the collective will and the social contract, which can create normative pressure to comply, as individuals may fear social or reputational costs for deviating from it. On the other hand, the act of voting itself can trigger an internal process of self-reflection and civic duty, promoting a personal

commitment to the common good and ultimately driving higher levels of cooperation and law compliance. This normative conundrum motivates our research questions: Does the expressive power of law depend on the visibility of the collective will, such that a widely publicized voting outcome exerts a stronger normative pull? Or is the very act of voting – irrespective of whether the result is made public – enough to trigger a process of introspection or self-nudging, thereby fostering compliance with the law?

To address these questions, we design a controlled laboratory experiment aimed at disentangling the behavioral mechanisms underlying the democracy premium in the context of expressive legal interventions. The experimental framework involves a modified stealing game that introduces a moral dilemma through the temptation to steal. Before getting the opportunity to take, participants receive symmetric endowments. In addition, we create an environment of normative ambiguity by pitting two competing norms: the norm prohibiting theft (*Thou shalt not steal!*) and the norm of merit (*You deserve what you worked for!*). Prior to the assignment of roles in the stealing game, participants compete in a real-effort task; the winner becomes the taker in the stealing game, and the loser becomes the victim, though our design is neutrally framed. Higher effort in the competitive real-effort task serves as a signal of competence and hard work, providing a moral basis for deserving additional rewards according to the norm of merit.[1] Higher effort or competence bolsters the argument that the taker is entitled to a larger share, intensifying the tension between the norm proscribing theft and the merit-based justification for taking. If takers prioritize the norm proscribing theft, the outcome is an egalitarian split equivalent to the experimenters' fiat endowment; if takers prioritize the norm of merit, a higher share of the pair's overall fiat endowment is assigned to the taker.

We implement three treatments to investigate the expressive power of democracy. The first is the No Voting (NV) treatment, in which participants simply play the stealing game without any democratic process. This condition serves as our baseline. In the second treatment, Hidden Result (HR), participants vote on whether stealing is socially acceptable but are not informed about the outcome of the vote, thereby primarily activating an introspective, reflective process. In contrast, the Revealed Result (RR) treatment makes the voting outcome publicly known, highlighting external social expectations and the salience of collective norms. While the HR treatment encourages self-nudging and behavioral self-management, the RR treatment more closely resembles a conventional democratic decision-making process in which the collective norm is publicly affirmed, potentially engendering a stronger sense of responsibility. To isolate the expressive effect of democracy, none of the treatments features external enforcement mechanisms. Overall, this experimental design is intended to clarify whether the expressive power of law depends on the salience of the voting result or whether the act of democratic participation itself can foster commitment to obey the law.

Our results reveal a nuanced interplay between normative ambiguity and the democracy

[1] We cannot exclude that takers might also experience a sense of status and power, although most participants refer to merit as the primary motive for taking in our post-experimental survey.

premium in expressive law. On average, takers steal slightly more than half of the victim's endowment. While theft is in line with rational choice theory, we find evidence that the norm of merit serves as justification to resolve normative ambiguity in a selfish manner, indicating that a strong perception of merit supports self-serving bias. However, the magnitude and likelihood of theft are lower in the RR treatment relative to the NV and HR treatments, highlighting the critical role of visible voting outcomes in shaping behavior. While the difference between the NV and HR treatments is insignificant, we do find that the RR treatment significantly reduces the magnitude and likelihood of theft compared to the baseline. These findings suggest that the democracy premium in expressive law can only be fully realized when the outcome of the vote is sufficiently salient, underscoring the importance of publicity in democratic processes.

Our study makes several contributions. First, it addresses a key gap in the literature on expressive law by advancing our understanding of the conditions under which the expressive effect of the law can be leveraged. Previous research has partly begun to explore these conditions, focusing on the impact of the content of the law (Nadler, 2024), of context (Teichman & Leibovitch, 2024), and of information aggregation (Dharmapala & McAdams, 2003). Our findings extend this literature by introducing a novel set of parameters – specifically, procedural factors – and demonstrating that the democratic process underlying the production of legal provisions can enhance compliance, even in settings characterized by normative ambiguity and a lack of formal or informal enforcement.

Second, we also contribute to the behavioral law and economics of stealing. Drawing on the stealing literature, we modify the stealing game to create a normatively ambiguous environment. Moreover, we document that the taking aversion reported in previous studies (Korenok et al., 2014, 2018) persists in normatively ambiguous contexts. We thus aim to uncover part of the cognitive and motivational forces driving theft and the role of communication through law as a potential means of deterrence.

Third, our study aligns with and extends the empirical literature on the democracy premium, which investigates how democratic processes enhance institutional effectiveness without modifying the severity or likelihood of punishment for norm violations (Dal Bó et al., 2010; Langenbach & Tausch, 2019; Marcin et al., 2019; Tontrup & Gaissmaier, 2023; Tyran & Feld, 2006). Existing studies on the democracy premium primarily focus on cooperative settings, exploring democracy in public goods games (Baldassarri & Grossman, 2011; Marcin et al., 2019; Tyran & Feld, 2006) and the prisoner's dilemma (Dal Bó et al., 2010). All these studies focus on the idea that democracy is mainly about shaping first- and second-order beliefs in interactive settings, leaving an open question: does the same effect persist in non-cooperative contexts? We extend the literature by exploring the democracy premium in a non-cooperative environment and comparing it with a purely introspective variant of democracy (i.e., voting without publicity of the outcome). Specifically, we aim at leveraging the power of democracy in a novel experimental context: the stealing game implemented under normative ambiguity.

Finally, our study makes a broader contribution to the law and economics literature on cost-efficient law enforcement. While enforcement is often considered a reliable mechanism for ensuring compliance, it is usually very costly. In contrast, communicating the outcomes of a vote is a relatively low-cost alternative. Our findings suggest that publicizing voting outcomes can partially substitute for enforcement by effectively promoting compliance with the law. This highlights the potential for leveraging democratic processes as a cost-efficient tool for enhancing legal adherence in comparison to costly enforcement.

The remainder of this paper proceeds as follows: Section 2 provides a review of the literature and explains our contribution. Section 3 outlines our experimental design. In Section 4, we develop our hypotheses. Section 5 presents the results of our study. We conclude with a discussion in Section 6.

2 Expressive Law, Stealing, and Democracy

Our paper connects to various strands of literature on expressive law, on the behavioral economics of stealing, and on the democracy premium. While expressive law defines the general mode of generating compliance with the law, we study theft as a potential object of deterrence, with democracy explored as the specific channel through which the law may exert expressive power.

The Expressive Power of Law. The concept of expressive law pertains to the capacity of legal rules to influence behavior independently of external enforcement mechanisms (Cooter, 1998; McAdams, 2015; Nadler, 2024; Sunstein, 1996). The literature identifies three potential mechanisms through which this effect operates.

First, the law serves as a coordination device, providing focal points in situations featuring the properties of coordination games (McAdams, 2000b). In such scenarios, two or more people benefit from making the same choice or aligning their behavior, even though multiple ways to coordinate may exist. Traffic laws offer a clear example: individuals gain from everyone driving on the same side of the road, whether left or right. A law dictating which side to drive on signals to individuals how they should behave to avoid accidents and thereby maximize collective benefits. Experimental and observational studies provide robust evidence for the expressive power of law and its underlying mechanisms. For example, both third-party suggestions (McAdams & Nadler, 2005) and legal suggestions (McAdams & Nadler, 2008) have been shown to facilitate coordination among individuals. Stressing the importance of communication in strategic interactions, legal scholars have argued that individuals may follow the law as a way to signal their character as law-abiding citizens to others (Posner, 1998, 2002).

Second, the law conveys information about acceptable behavior within a society (McAdams, 2000a). People often adjust their behavior to align with societal expectations (Cialdini et al.,

1990). Laws act as an effective medium for communicating what the majority perceives as appropriate conduct in a given context. Applying the Condorcet Jury Theorem, Dharmapala and McAdams (2003) argue that the legislative process can aggregate crucial information about the world, thereby causing citizens to update their prior beliefs and change their behavior. When a law signals expectations about a specific type of behavior, individuals update their beliefs and align their actions accordingly. For example, a non-binding request for a minimum contribution in a public goods game significantly increases the level of contributions (Galbiati & Vertova, 2014). Similarly, framing specific performance as the default norm leads to higher spending to avoid an efficient breach compared to expectation damages as the default rule (Depoorter & Tontrup, 2012). Recent studies exploring behaviors that are legal or illegal depending on a specific threshold, such as consuming alcohol or driving within speed limits show that the discontinuity introduced by the law at these thresholds – distinct from social norms – is mirrored in individuals' perceptions of the behaviors in question (Görges et al., 2023; Lane et al., 2023; Teichman & Leibovitch, 2024).

Third, expressive law scholars argue that compliance with the law is often shaped by a meta-norm of legal obedience or civic duty (McAdams & Rasmusen, 2007). To avoid the symbolic or internal psychological costs of violating this meta-norm, individuals tend to comply with laws even in the absence of effective sanctions. Exploring data from Switzerland, Funk (2007), for example, shows that the abolition of a voting duty subject to a purely symbolic fine significantly decreased average turnout. Stressing the moral dimension of the law, engineers in Silicon Valley report that their compliance with laws against sharing trade secrets is motivated by personal moral principles (Feldman, 2006), though this effect does not extend to the context of digital file sharing (Feldman & Nadler, 2006). Furthermore, during the COVID-19 pandemic, laws reinforced the belief that social distancing measures were necessary and should be followed (Galbiati et al., 2021). Interestingly, individuals who updated their behavior in response to pandemic regulations reverted their behavior after the removal of those laws (Casoria et al., 2021).

The central claim of the expressive law literature – that law can influence behavior without coercion – has not remained uncontested (Schauer, 2015). This is partly due to the scarcity of empirical studies investigating the specific conditions under which the expressive power of law can be leveraged (cf. Nadler, 2017, 2024; Teichman & Leibovitch, 2024). Our study aims at partly filling this gap by exploring the interplay between information about group behavior, legitimacy, and normative ambiguity. First, when a law conveys information about a norm that conflicts with the social identity of a particular group, it fails to affect the members of that group (Nadler, 2017, 2024). Second, in normatively ambiguous contexts, individuals are more likely to seek legal guidance to inform their own behavior or evaluate the behavior of others (Teichman & Leibovitch, 2024). Third, while the effects of procedure and legitimacy on compliance are well-documented in other areas, for example in procedural justice research (Tyler, 2006, 2010), its role in shaping the expressive effect of the law remains unclear.

The Morality of Stealing. Rational choice theory assumes selfish choices. In Gary Becker's model of crime and punishment the decision to commit a crime is based on the comparison of expected utility from committing the crime versus not committing it (Becker, 1968). Formally, a criminal will maximize the expected utility from committing a crime:

$$EU_{\text{crime}} = (1 - p) \cdot U(W + G) + p \cdot U(W - F),$$

where U denotes the utility of the individual, W individual wealth if no crime is committed, G the gain from committing the crime, p the probability of being punished, and F the monetary equivalent of the punishment.

The Beckerian model of crime thus predicts that individuals will fully expropriate others' property in the absence of a threat of punishment. This prediction is tested experimentally using the stealing game – sometimes referred to as the taking game (for a general overview, see Flage, 2024). Corroborating the Beckerian model of crime and punishment, numerous experimental studies show that the presence of punishment is a key determinant in reducing stealing behavior (Engel, 2016; Engel & Nagin, 2015; Khadjavi & Lange, 2015; Rizzolli & Stanca, 2012). For example, research has explored the interaction between deterrence and other factors, such as risk preferences explaining the importance of punishment certainty (Engel & Nagin, 2015) and the crowding-out effect of deterrence on prosocial emotions (Khadjavi & Lange, 2015).

However, a growing body of literature challenges the assumption that the absence of punishment leads to full expropriation (Gravert, 2013; Korenok et al., 2014, 2018; List, 2007). These studies posit that moral costs – the psychological costs associated with taking – can explain why individuals refrain from taking the entirety of their matched partner's endowment. Moral costs include personal preferences against taking (Korenok et al., 2014, 2018) and internal punishments, such as feelings of shame or guilt for violating a social norm (Krupka & Weber, 2013). Previous research shows how participants receive the endowment (earned/windfall) has significant effect on decision about amount to be stolen (Faillo et al., 2019; Gravert, 2013; List, 2007). These findings highlight the possibility to create a normative conflict, by combining a temptation to steal – backed by the sense of merit – with a vote questioning the moral admissibility of stealing. Taken together, these findings suggest broad support of a social norm prohibiting theft.

The discrepancy between the predictions of rational choice theory and experimental findings indicating varying degrees of altruism or prosocial behavior has prompted deeper inquiry into the cognitive and motivational forces shaping fair behavior (Henrich et al., 2004). A strand of research in experimental moral philosophy offers compelling evidence that individuals exploit situations where there is an obscure causal link between their unfair actions and unfair outcomes, a phenomenon referred to as moral wiggle room (Bosco, 2022; Dana et al., 2006, 2007).

In the context of the dictator game, for example, participants are more likely to select an unfair outcome when they are uncertain whether their actions or a coin toss determines the result. Notably, they tend to avoid acquiring information that would clarify this uncertainty. Scholars posit that acting unfairly comes at a psychological cost due to the misalignment between one's actions and self-perception (Bosco, 2022). When individuals have the opportunity to behave selfishly without incurring this psychological cost, they tend to opt for unfair outcomes. In a seminal experiment exploring the effects of moral wiggle room, Dana et al. (2007) implement four treatments of the dictator game, a common tool for measuring altruism. The baseline treatment features complete transparency, while other treatments involve opaque conditions, where key information linking behavior to outcomes was missing. The authors find that participants are more likely to choose unfair outcomes under opaque conditions and actively avoid acquiring missing information.

These and related findings have been corroborated by other studies (Larson & Capra, 2009). More generally, the implementation of interactive games (van der Weele et al., 2014), manipulating the type of missing information (Thunström et al., 2016), varying the risk of being publicly identifiable as the originator of the decision in a dictator game (Hermstrüwer & Dickert, 2017) have been shown to influence the extent of unfair behavior.

The Democracy Premium. Previous studies demonstrate that democratically chosen rules and institutions are generally more effective than those imposed exogenously (Sutter et al., 2010; Tyran & Feld, 2006). This phenomenon extends beyond the self-selection of cooperative individuals into cooperative institutional setups (Dal Bó, 2014; Dal Bó et al., 2010) and includes the direct impact of democratic processes on rule or institution acceptance. Exploring a public goods game with internal centralized sanctioning authorities, Grossman and Baldassarri (2012), for example, show that the level of cooperation is contingent on the political process by which the authority originally acquires its sanctioning powers. Similar effects have been predominantly studied in cooperative contexts, such as public goods games (Baldassarri & Grossman, 2011; Kamei, 2016; Langenbach & Tausch, 2019; Marcin et al., 2019; Sutter et al., 2010; Tontrup & Gaissmaier, 2023) and the prisoner's dilemma (Dal Bó et al., 2010, 2019), and across various institutional frameworks, including minimum contribution requirements and punishment mechanisms.

Empirical research has uncovered intriguing pathways in which democratic processes influence cooperation and rule-following more broadly. Specifically, studies have found that individuals display higher levels of cooperation when institutions are established through a democratic vote rather than when they are imposed exogenously (Dal Bó et al., 2010). Moreover, when third-party enforcers are chosen democratically, they tend to impose less severe sanctions compared to those implemented exogenously (Marcin et al., 2019). Additionally, granting participation and voting rights appears to bolster cooperative behavior (Tontrup & Gaissmaier, 2023). While the precise mechanisms driving the enhanced effectiveness of demo-

cratically selected institutions over exogenously imposed ones remain unclear, three main explanations have emerged: self-selection, the signaling of cooperativeness, and the proposition that the democratic process itself can shape behavior regardless of self-selection and signaling – an idea that this study explores further.

3 Experimental Design

Basic Setup. The basic setup is the same across all three treatments and proceeds in two main steps. Participants act under complete information about the entire experimental protocol throughout the experiment. In the first main step, participants are randomly matched in pairs and then take part in a simple (competitive) real-effort task. The task consists in counting the number of zeroes in a series of tables containing numbers for a duration of five minutes. The participant solving more tables is assigned to the role of taker, the other participant is assigned to the role of victim, with ties in the number of tables solved being broken randomly. This design feature is aimed at inducing a sense of merit for the taker. While winning the real-effort competition does not directly create a right to steal, it is intended to trigger a perception of merit and self-serving bias, making it easier for winning participants to justify selfish behavior or an outcome favoring her over losing participants. And our manipulation check indeed shows that participants solving more tables in the real-effort task indeed experienced a sense of merit.[2] To avoid potential framing effects, we deliberately opted for a neutral wording and referred to Player A (the taker) and Player B (the victim) throughout the experiment.

In the second main step, participants take part in a stealing game with symmetric endowments, where each taker and each victim is endowed with 10 points. Player A (the taker) is told that they are allowed to take any amount from Player B (the victim) they have been paired with. Participants are told that this is a one-shot game. Before proceeding to the stealing game, we elicit takers' beliefs about their relative performance in the real-effort task, allowing us to assess whether and how performance beliefs explain taking. Each participant is asked to assess the numerical distance in the number of tables solved between themselves and the participant they are paired with. Participants finish the experiment by answering to a socio-demographic questionnaire.

Treatments. We design three treatments: a No Voting treatment (NV) that proceeds exactly according to the protocol of the aforementioned basic setup and serves as our baseline treatment, a Hidden Result treatment (HR), and a Revealed Result treatment (RR). In the HR and RR treatments, all participants are assigned to groups of six, comprising three pairs of partic-

[2]Specifically, participants indicated that they took points from the player they were matched with because they considered the allocation unfair. We interpret these statements as evidence that our design was successful in creating normative ambiguity associated with competing applicable norms. See Appendix A.3 for a summary of the explanations and arguments that participants made in support of their behavior in the stealing game and their voting behavior.

ipants that are randomly matched for the real-effort task in the first main step. Each group is designed as a matching group, meaning that it corresponds to one independent observation in our econometric analysis. Participants in each group take part in a voting procedure implemented after the real-effort task and after the belief elicitation stage but before the disclosure of roles, a design feature that is Rawlsian in spirit. Having participants decide behind a veil of ignorance (cf. Rawls, 1971) is intended to make it more likely for a truthful and unbiased vote to emerge. Participants cast a vote on the following norm and cannot refuse to vote:

Person A is allowed to take what they deserve from Person B.

While neutrally framed regarding the specific roles, the question is framed to prompt participants to consider the norm of merit. The explicit reference to merit is intended to make it easier for participants to identify a normative justification for behaving selfishly and inducing the perception of a normative conflict between the norm of merits (selfish motives) and the norms proscribing theft (altruistic motives). Participants cast their vote by clicking either *YES* or *NO*.

The HR treatment is designed to activate an introspective process, gently nudging participants to consider the appropriate course of action. To achieve this, participants do not receive any information about the outcome of the vote. They are merely asked to cast their vote, with the outcome of the vote kept secret – an unobtrusive and privacy-preserving way of implementing the democratic vote and bringing the norm against theft into focus. After having cast their vote they proceed to the stealing game.

The RR treatment proceeds exactly as the HR treatment, but for one difference: unlike in the HR treatment, participants are informed about the outcome of vote in their group – the "micro-society" consisting of six participants. The outcome of the vote becomes public before participants proceed to the stealing game. This design feature is intended to mimic the public communication of the outcome obtained in the democratic process. The outcome is presented neutrally to avoid any framing effects, with participants being informed how many peers in their group voted in favor of the norm of merit justifying theft.

Procedure. A total of 492 participants were recruited across 26 sessions held between July and November 2023 at the Decision Lab of the Max Planck Institute for Research on Collective Goods. The final sample includes 480 observations, with 240 participants assigned the role of taker and 240 assigned the role of victim.[3] 120 participants were assigned to the NV and to the

[3]Prior to the lab experiment we also conducted an online experiment (see Appendix A.2) using the incentivized norm elicitation method proposed by Krupka and Weber (2013). Socially appropriate behavior was elicited using two vignettes describing two window cleaners, John and Bob, receiving the same amount of money at the end of the day. The vignettes differed in the described performance of the window cleaners. While both window cleaners cleaned the same number of windows in the first vignette, John cleaned more windows than Bob in the second vignette. Participants were told that John sees the envelope with Bob's salary on a desk and is tempted to steal part of the money. In both cases, most participants agreed that refraining from theft was the most socially appropriate action. While people seem to be highly aware of the social norm proscribing theft,

HR treatment respectively; 240 were assigned to the RR treatment. This numerical imbalance was necessary to account for the fact that participants in the RR could see the voting outcome in their group before proceeding to the stealing game, resulting in the dependence of stealing decisions. Given that participants were assigned to matching groups of six consisting of three taker-victim pairs, we generate a sample of 40 independent observations in the RR treatment.

4 Hypotheses

The Beckerian model predicts that a taker – not exposed to any risk of formal or informal punishment – steals all points from the victim. However, as discussed above, behavioral forces will often induce people to refrain from stealing even when doing so is rational. This discrepancy between theoretical predictions and observed behavior arises from the moral costs associated with stealing (Gravert, 2013; Levitt & List, 2007). Normative ambiguity creates moral wiggle room, diminishing the moral costs of engaging in theft. Participants, aiming to preserve their self-image as moral individuals can be expected to justify immoral actions by invoking the norm of merit.

Our design exploits three behavioral forces that we expect to shape behavior. First, decision making under ambiguity allows individuals to preserve their self-image as moral persons while engaging in unethical behavior (Dana et al., 2007). Participants in treatments with minimal information and the highest levels of ambiguity, the NV and the HR treatments, are expected to behave more selfishly, exploiting moral wiggle room. Second, the salience of a norm plays a critical role in determining its behavioral impact (Stok & de Ridder, 2019). The framing of our voting procedure in the HR and RR treatments is intended to draw participants' attention to the right to take, thereby increasing the salience of the conflict between the norm of merit and the conventional prohibition of theft. We expect this to induce participants to actively engage with the tension between the two competing norms. Third, when a norm is not only prominent but also visibly endorsed by the collective and perceived as legitimate, any deviation is experienced as a moral transgression, thereby discouraging norm violations. Violating such a norm should come at a greater moral cost in the RR treatment than in the HR treatment. Based on these mechanisms, we propose the following hypothesis:

- *H1: The probability and magnitude of stealing are higher in the NV treatment compared to the HR and RR treatments.*

Participants are unaware of how well they perform relative to their counterparts in the real-effort competition. However, their beliefs about relative performance serve as a justification mechanism, reducing the moral costs of stealing. Individuals inclined to take more from their

participants in our lab experiment were much more inclined to express support for the conflicting norm of merit justifying theft. We control for participation in the online experiment in our regression estimates of treatment effects observed in the lab experiment.

partner may adjust their beliefs to align selfish behavior with their moral self-image. Thus, we propose:

- *H2: The belief about relative performance in the real-effort competition positively correlates with the probability and magnitude of stealing.*

Endorsing a social norm while simultaneously violating it is in line with the predictions of rational choice theory. This potential ambiguity between social norm endorsement and actual behavior limits the ability to disentangle the effects of voting from self-interest motives in our estimation models. However, if participants reject the norm and act accordingly, this indicates that individuals seek consistency between their attitudes and behavior, even if this entails a monetary cost. Hence, their behavior can be attributed to the effect of voting. Therefore, we expect:

- *H3: The probability and magnitude of stealing are lower among participants who reject the norm compared to those who approve it.*

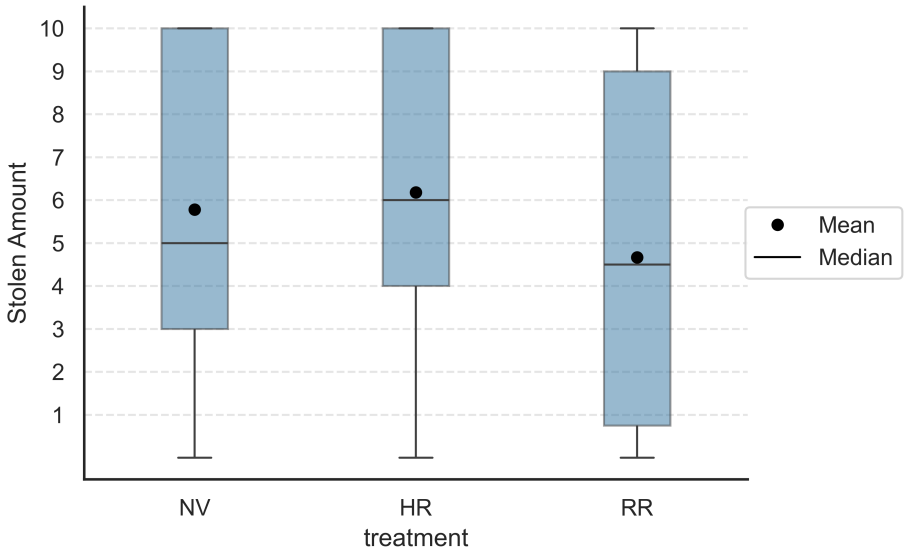
5 Results

We begin by presenting the results on our treatment effects (5.1), followed by an examination of heterogeneous treatment effects (5.2). We then analyze the effects of voting (5.3), and offer an exploratory investigation into participants' propensity to exploit moral wiggle room under conditions of normative ambiguity (5.4).

5.1 Treatment Effects

Main Effects. We begin our analysis by comparing the average amount stolen across treatments using a two-sided Mann-Whitney U test. While the difference between the NV (No Voting: $m = 5.78$, $SD = 3.69$) and HR (Hidden Result: $m = 6.18$, $SD = 3.64$) treatments is not statistically significant ($p = 0.485$), we find that the RR treatment (Revealed Result: $m = 4.67$, $SD = 3.89$) reduces stealing relative to the NV treatment. This effect is marginally significant ($p = 0.069$). A similar but stronger effect emerges when comparing the HR treatment and the RR treatment ($p = 0.01$). This result, visualized in Fig. 1, lends partial support to hypothesis *H1*. A summary of the average amount stolen and of the Mann-Whitney U test can be found in the Appendix (Tables 6 and 7 in Appendix A.1).

Figure 1: Amount stolen across treatments



Investigating the proportion of participants stealing at least something or nothing, we obtain similar results: the proportion of participants who steal at least something is similar in the NV and HR treatments, but it is significantly lower in the RR treatment (see Fig. 6 in Appendix A.1). While 87% of takers in the NV and HR treatments steal something, this proportion drops to 75% in the RR treatment. These results lend support to the conclusion that the expressive power of democracy is highly contingent on the public visibility of the normative support it generates.

In addition to the non-parametric tests, we also use parametric tests, estimating various specifications of an OLS regression model for the amount stolen Y_i with the following full specification:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 (x_{1i} \times x_{2i}) + \beta_4 x_{3i} + \beta_5 \phi_i + \varepsilon_i,$$

where x_{1i} denotes treatment dummies; x_{2i} denotes voting behavior, coded as 0 for rejection and 1 for approval; x_{3i} denotes participants' beliefs about their relative performance compared to their counterparts, measured by the difference in the number of tables solved. We include the interaction term $x_{1i} \times x_{2i}$ to explore the moderating effect of voting behavior on the relationship between treatment and stealing. ϕ_i is a vector of sociodemographic control variables, including age, gender, employment status at the time of the experiment, student status, participation in our online norm elicitation experiment (see Appendix A.2), prior participation in experiments, and the number of incorrect answers to control questions. The error term ε_i accounts for unobserved factors affecting the outcome.

In the simplest specification, we regress the amount stolen on two treatment dummies (Table 1, Model 1). Its results indicate that the RR treatment yields a marginally significant reduction in stealing as compared to the NV treatment. While the HR treatment nominally increases

the amount stolen relative to the NV treatment, this effect is statistically insignificant. Estimates of a linear mixed model yield similar results (see Table 8 in Appendix A.1).

Table 1: Treatment effects

DV: Amount stolen	(1)	(2)	(3)
Revealed Result	-1.117*	-1.069*	-1.008*
	(0.604)	(0.582)	(0.574)
Hidden Result	0.400	0.388	0.255
	(0.706)	(0.748)	(0.766)
Performance Beliefs			0.854***
			(0.293)
Sociodemographics		✓	✓
Constant	5.783***	6.898***	6.194***
	(0.513)	(1.613)	(1.668)
Observations	240	240	240
R^2	0.031	0.119	0.147
Adjusted R^2	0.023	0.072	0.098

This table reports the results of an OLS regression model. Standard errors in parentheses and clustered at the group level. The *No Voting* treatment serves as the reference category in all model specifications. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

These results remain robust when controlling for beliefs and sociodemographics (Table 1, Model 2 and 3). Including these controls slightly reduces the magnitude of our main treatment effects, but the coefficient of the RR treatment remains significant. Using a logistic regression model for the probability of theft, we obtain similar results (Table 10 in Appendix A.1). In sum, these results corroborate the findings obtained using the non-parametric tests: the RR treatment significantly reduces stealing, but the HR treatment yields no statistically significant decrease in stealing.

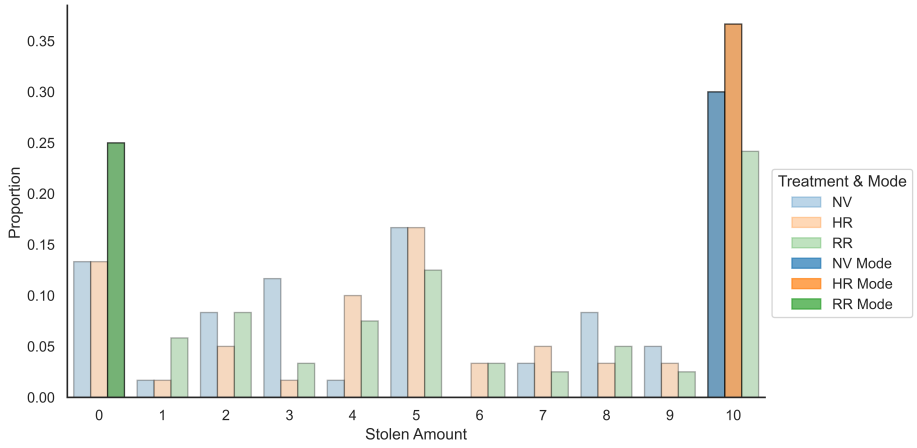
- *Result 1: The RR treatment reduces stealing relative to the NV treatment.*

Distributive Effects. A detailed examination of the distribution of stealing reveals a striking contrast between the RR treatment and the other treatments (Fig. 2). Our analysis reveals an almost trimodal distribution, with distinct peaks at 0 (none), 5 (half), and 10 (all). Notably, 61% of participants choose one of these three amounts, with 19% stealing nothing, 14% taking half (5 points), and 28% taking all (10 points). In the RR treatment, most participants take nothing, whereas most participants take the victim's entire endowment in the NV and HR treatments.

This suggests that the RR treatment entails a substantial behavioral shift towards the norm proscribing theft. The NV and HR treatments, by contrast, seem to push participants to steal the entirety of the victim's endowment. Corroborating this observation, we find that the distributions marginally differ when comparing the HR and RR treatments (Kolmogorov-Smirnov

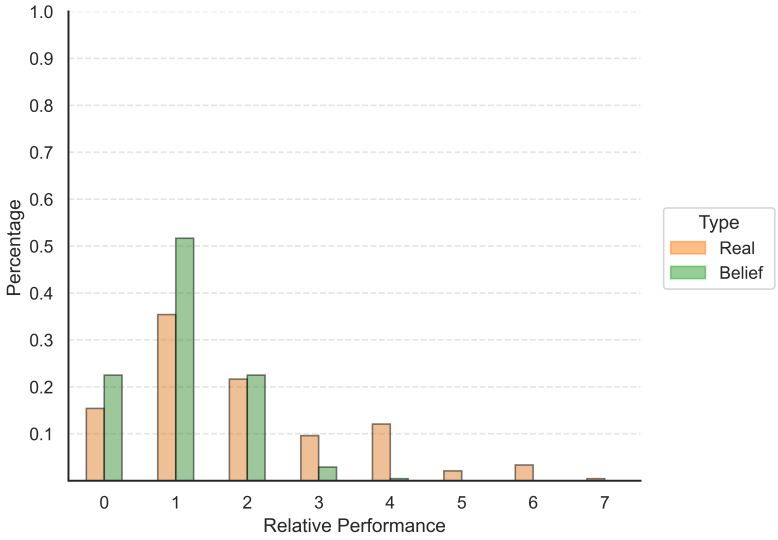
test, $D = 0.208$, $p = 0.059$).[4] This shift in the distribution indicates that participants exploit normative ambiguity to their benefit whenever the norm proscribing theft is not communicated at all or when the vote about the socially appropriate course of action remains a merely introspective act.

Figure 2: Distribution of amount stolen



Impact of Beliefs. Many people strive to perceive themselves as moral, even when engaging in immoral actions. To reconcile this dissonance, they rationalize their unethical behavior when given the opportunity. In our experiment, participants can justify their stealing behavior by referencing their beliefs about relative performance. The distribution of participants' beliefs regarding their relative performance and their actual relative performance is visualized in Fig. 3. Notably, many participants tend to slightly overestimate their performance in the real-effort competition. These self-serving beliefs fuel the propensity to steal.

Figure 3: Proportion of beliefs and real performance



[4] A comparison between the NV and RR treatments reveals no significant difference in distributions ($D = 0.158$, $p = 0.260$); neither does the comparison between NV and HR treatments ($D = 0.133$, $p = 0.665$).

Corroborating the effect of self-serving beliefs, estimates of our OLS regression indicate a positive correlation between beliefs and stealing (Table 1, Model 3). For each additional table solved according to perceived relative performance, the amount stolen increases by 0.854 points – a highly significant effect. These findings are consistent with the idea that individuals tend to exploit opportunities to justify immoral behavior, and lend support to hypothesis *H2*.

- *Result 2: The amount stolen increases with the belief about relative performance in the real-effort competition.*

5.2 Heterogeneous Treatment Effects

A natural question emerging from our analysis of distributive effects is whether the effect of the RR treatment differs between individuals with high and low moral costs of stealing (i.e., between individuals who are inclined to engage in petty theft and those who are prone to heist). The moral costs of stealing are likely to vary among individuals and might explain differences in theft. To explore this question, we estimate two specifications of a quantile regression model across different quantiles of the stealing distribution, specifically the 20th, 50th, and 80th percentiles (Table 2). This analysis rests on the assumption that participants in lower quantiles of the stealing distribution incur higher moral costs of theft (i.e., they would steal more otherwise) than participants in upper quantiles of the distribution (i.e., they would steal less otherwise).

We find that the RR treatment significantly reduces stealing relative to the NV treatment at the lower quantile (20th percentile), while we observe no significant treatment effect at higher quantiles (Table 2, Model 1). Corroborating our previous results, we again find no significant effect of the HR treatment compared to the NV treatment. Overall, this result suggests that the RR treatment is most effective for participants who incur relatively high moral costs of theft, while participants who incur lower moral costs of theft remain, by and large, unaffected by the RR treatment. This finding points to an important limitation of the expressive power of democracy and of law more generally: while publicly conveying information about norms may effectively curb the behavior of those inclined toward minor norm violations, it seems much less effective in reigning in those already strongly predisposed to break the law.

Table 2: Treatment effects across quantiles

	(1)			(2)		
DV: Amount stolen	20%	50%	80%	20%	50%	80%
Revealed Result	-1.402**	-0.883	0.008	-2.137**	-0.730	-0.107
	(0.673)	(0.849)	(0.659)	(0.985)	(0.926)	(0.911)
Hidden Result	-0.106	0.837	0.093			
	(0.806)	(0.971)	(0.742)			
Norm Rejected				-4.373***	-3.689***	-2.821***
				(1.261)	(1.103)	(0.983)
Revealed Result x Norm Rejected				2.392	-0.066	-0.903
				(1.450)	(1.345)	(1.297)
Performance Belief	1.052***	1.280***	0.042	0.608	0.716	0.010
	(0.334)	(0.448)	(0.326)	(0.369)	(0.440)	(0.489)
Sociodemographics	✓	✓	✓	✓	✓	✓
Constant	0.712	5.984***	10.279***	4.608**	8.623***	10.004***
	(1.881)	(2.147)	(1.646)	(1.931)	(1.991)	(1.832)
Observations	240	240	240	180	180	180
Pseudo R^2	0.101	0.105	0.048	0.168	0.220	0.139

This table reports the results of a quantile regression model. Standard errors in parentheses. The *No Voting* treatment serves as the reference category in Model 1. Those who approve the norm in the *Hidden Result* treatment serve as the reference category in Model 2. Model 2 does not include the *No Voting* treatment. The dependent variable is in continuous format.
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

We obtain a similar result when estimating the effect of the RR treatment relative to the HR treatment in a quantile regression model that includes a dummy variable for norm rejection and an interaction term between the RR treatment and norm rejection (Table 2, Model 2). The inclusion of these covariates restricts the reference category to participants who approve the norm in the HR treatment. For participants who incur higher moral costs of theft, i.e., participants at the 20th percentile, the RR treatment reduces theft by 2.137 points (approximately 21% of the endowment) relative to the HR treatment. The effect is insignificant for those with lower moral costs of stealing (i.e., participants at the 50th and 80th percentile).

The main effect of norm rejection reflects the behavior of participants who reject the norm in the RR treatment, relative to the HR treatment. Rejecting the norm of merit significantly reduces stealing in all quantiles; however, the magnitude of the effect varies. Specifically, the reduction is 1.5 times greater in the 20th percentile than in the 80th percentile. This further corroborates that rejecting the norm of merit has a stronger effect on individuals with higher moral costs of stealing than on those with lower moral costs. More broadly, the results support the conclusion that withholding the outcome of the voting process may create moral wiggle room, particularly for individuals with stronger selfish tendencies. The interaction term compares individuals who reject the norm of merit in the RR treatment with those in the HR treatment. We find no significant behavioral differences between these two groups. Publicizing the outcome of the vote thus produces similar patterns for norm-approvers and norm-rejecters alike.

These findings suggest that people who steal less and approve the norm of merit are likely

to experience greater internal conflict. They adhere to the norm but simultaneously wish to change it. Public disclosure of the voting outcome influences their behavior. Taken together, these quantile results suggest that the RR treatment primarily influences those with high moral costs of theft, while both norm-approvers and -rejecters at higher quantiles remain unaffected.

5.3 Voting

Endogenous Effects: Choosing Norms. One particularly intriguing question is whether rejecting the norm of merit signals a normative commitment to refrain from stealing. As Table 3 shows, voting behavior is relatively consistent across treatments when pooling both takers and victims. However, we observe quite some variation when only analyzing takers. Approval of the norm of merit is substantially higher in the HR treatment, which suggests that participants – before being assigned the role of taker and thus behind a veil of ignorance – feel tempted to express selfish motives, anticipating that their vote will never become public whatsoever.

Table 3: Voting behavior across treatments

Treatment	Voters	Approve	Reject
All	Takers and Victims	0.44	0.56
Hidden Result Treatment (HR)	Takers and Victims	0.46	0.54
Revealed Result Treatment (RR)	Takers and Victims	0.43	0.57
All	Takers	0.49	0.51
Hidden Result Treatment (HR)	Takers	0.57	0.43
Revealed Result Treatment (RR)	Takers	0.48	0.52

Estimates of an OLS regression confirm that the deliberate choice to reject the norm of merit signals a commitment to refrain from stealing. The effects of the RR treatment and of the relative performance belief are statistically significant (Table 4, Model 1). However, when controlling for norm rejection, the coefficient for relative performance beliefs becomes insignificant (Table 4, Model 2), suggesting an interdependence between these factors. Furthermore, the treatment effect on voting behavior does not depend on norm rejection, as the interaction term between treatment and norm rejection is insignificant (Table 4, Model 3).

Table 4: Effects of revealed results and norm rejection

DV: Amount stolen	(1)	(2)	(3)
Revealed Result	-1.198*	-1.054*	-1.182*
	(0.670)	(0.603)	(0.644)
Norm Rejected		-2.870***	-3.051**
		(0.583)	(1.219)
Revealed Result x Norm Rejected			0.275
			(1.343)
Performance Belief	0.864**	0.424	0.430
	(0.372)	(0.355)	(0.360)
Sociodemographics	✓	✓	✓
Constant	5.924***	7.770***	7.878***
	(1.871)	(1.691)	(1.622)
Observations	180	180	180
R^2	0.162	0.282	0.282
Adjusted R^2	0.102	0.226	0.221

This table reports the results of an OLS regression model. Standard errors in parentheses and clustered at the group level. Those who approve the norm in the *Hidden Result* treatment serve as the reference category; estimations do not include participants in the *No Voting* treatment.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A mediation analysis indicates that norm rejection mediates the relationship between the performance belief and stealing (Table 9 in Appendix A.1). Participants who perceive a smaller performance gap relative to their counterpart tend to reject the norm, resulting in reduced theft. Furthermore, rejecting the norm independently and significantly decreases the amount stolen, regardless of performance beliefs. The direct effect of the performance belief on theft is insignificant; however, its indirect effect through norm rejection is highly significant after bootstrapping confidence intervals, suggesting that the performance belief influences stealing behavior exclusively via changes in voting behavior. Importantly, the negative effect of the RR treatment on the amount stolen, compared to the HR treatment, remains marginally significant. Together, these results support *H3*: rejecting the norm of merit is associated with a significant decrease in theft by altering beliefs about relative performance in the real-effort competition.

- *Result 3: Rejecting the norm of merit is correlated with a reduction in the amount stolen.*

Exogenous Effects: Observing Norms. A related question is whether the reduction in stealing observed in the RR treatment may be attributed to the information participants receive about the outcome of the voting process and to the strength of observed support for the norm of merit. The outcome obtained in each experimental "micro-society" conveys explicit information about the prevailing social norm, which participants may in turn choose to follow. In our experiment, possible outcomes range from 0 to 6 votes in favor of the norm of merit. Most participants observe that either two or three group members voted in favor of this norm, while

they never observe a consensual vote with every group member approving or rejecting the norm of merit (see Fig. 7 in Appendix A.1).

To investigate the effect of observing norms we begin by constructing meaningful reference categories, the first describing whether the norm of merit receives support from a minority – one vote or two votes – (RR: Minority vote), and the second describing whether the norm of merit receives one vote (RR: 1 vote). We find that observing outcomes of three or more votes does not significantly affect stealing compared to observing a minority vote (Table 5, Model 1). When comparing stealing among participants who observe one vote with participants who observe stronger support for the norm of merit, we find that observing two votes reduces theft by 1.625 points (approximately 16% of the endowment) relative to observing just one vote; yet more support for the norm of merit does not entail any significant reduction in stealing (Table 5, Model 2). In sum, we find little evidence that exogenous information about the prevailing norm affects stealing.

Table 5: Effect of observing votes

DV: Amount stolen	(1)	(2)
Revealed Result: 50-50 vote	0.383	
	(0.711)	
Revealed Result: Majority vote	0.375	
	(0.832)	
Revealed Result: 2 votes		-1.625**
		(0.765)
Revealed Result: 3 votes		-0.780
		(0.802)
Revealed Result: 4 votes		-0.958
		(0.872)
Revealed Result: 5 votes		-0.586
		(0.983)
Performance Belief	0.774*	0.844*
	(0.432)	(0.435)
Norm Rejected	-2.621***	-2.700***
	(0.645)	(0.677)
Sociodemographics	✓	✓
Constant	5.217**	6.157***
	(2.097)	(2.152)
Observations	120	120
R^2	0.313	0.331
Adjusted R^2	0.221	0.227

This table reports the results of an OLS regression model. Standard errors in parentheses and clustered at the group level. In model 1, *Revealed Result: Minority vote* serves as the reference category to compare the stolen amount between those who observe the minority outcome and those who observe the majority or equal outcome. In model 2, *Revealed Result: 1 vote* is the reference category, allowing us to compare the stolen amount between those who observe the minimum number of votes and those who observe other vote totals.

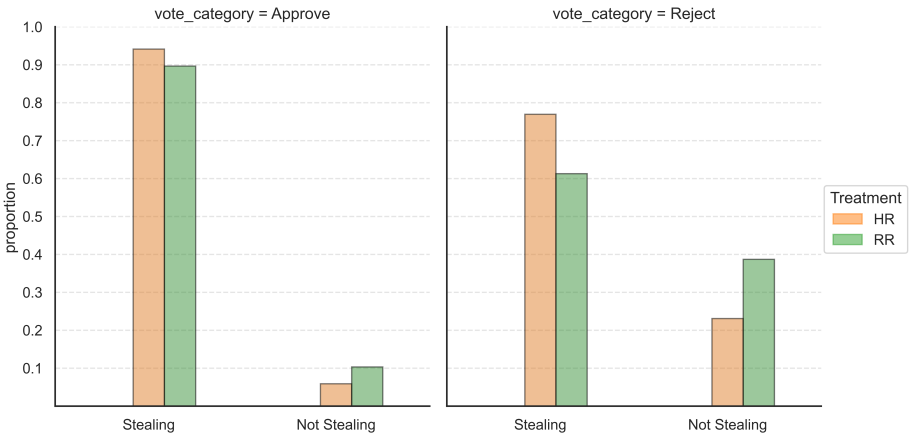
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.4 Moral Wiggle Room

Moral wiggle room refers to settings in which individuals can act immorally while preserving their self-image as moral agents. In experimental settings, moral wiggle room is typically created by obscuring information that links choices to outcomes, so that participants can plausibly deny responsibility for unfair outcomes (Dana et al., 2007). Our experiment is designed to explore a more active channel: rather than withholding information, selfish participants can reject the norm of merit and thereby convince themselves they have upheld and adhered to the norm proscribing theft.[5]

Figure 4 plots the share of participants who steal, separately by whether they approved (left) or rejected (right) the merit norm, and by treatment. Two patterns stand out. First, those who reject the merit norm steal less overall than those who approve it. Second – and more interestingly – participants rejecting the norm of merit are more likely to steal in the HR treatment than in the RR treatment. This observation suggests that the HR treatment makes it easier to reconcile *expressed preferences* against theft and *revealed preferences* in favor of theft.

Figure 4: Probability of stealing by voting and treatments



Nonparametric tests confirm these descriptive moral wiggle room patterns. We employ a two-sided Mann-Whitney U test to compare the amount stolen between participants who reject the norm ($m = 3.48$, $SD = 3.47$) and those who approve it ($m = 6.78$, $SD = 3.54$). The difference is highly significant ($p < 0.001$). Furthermore, participants who reject the norm in the HR treatment steal marginally more ($m = 4.42$, $SD = 3.46$) than those in the RR treatment ($m = 3.09$, $SD = 3.42$, $p = 0.097$). Similarly, for those who approve the norm, the difference between the HR treatment ($m = 7.52$, $SD = 3.21$) and the RR treatment ($m = 6.34$, $SD = 3.68$) is also marginally significant ($p = 0.093$). These results indicate that participants anticipating a private vote feel tempted to exploit moral wiggle room by endorsing the prohibition on theft to preserve a positive self-image, while opting to steal nonetheless.

[5]Another, related source of moral wiggle room is the conflict between the norm proscribing theft and the norm of merit. Allowing participants to express altruistic motives by rejecting the norm proscribing theft makes it easier to exploit moral wiggle room when facing the two conflicting norms.

6 Discussion

In this study, we investigate whether democratic voting procedures can enhance the expressive power of legal norms under conditions of normative ambiguity. Specifically, participants play a stealing game that pits two conflicting norms: the prohibition of theft (*Thou shalt not steal!*) versus the norm of merit (*You should get what you deserve!*). We introduce two voting mechanisms designed to elicit or reinforce the socially appropriate course of action. By comparing three treatments – the No Voting (NV), the Hidden Result (HR), and the Revealed Result (RR) treatments – we aim to disentangle two potential pathways by which democracy may boost compliance with the law: (i) an introspective deliberation triggered by voting, and (ii) publicly visible information about the collective will.

Findings. We find that the RR treatment yields a significant reduction in stealing compared to the NV baseline, suggesting that publicly visible support for a social norm can effectively steer behavior in the face of normative conflicts. In contrast, the HR treatment, which conceals the voting outcome, does not mitigate theft. This finding points to a democracy premium in expressive law that hinges on publicizing the collective will.

The context that our participants interact in is inherently ambiguous, and our manipulation checks corroborate that our design operates as intended: participants clearly experience normative ambiguity arising from the conflict between the prohibition against theft and the norm of merit justifying theft. As a result, participants in our experiment are, on average, neither entirely selfish nor entirely altruistic. However, we do observe quite some weight on the extremes of our distribution, indicating that many people opt for an all-or-nothing strategy when facing normative ambiguity.

Contrary to our expectation that a private, introspective voting process might induce participants to reflect on and abide by the norm proscribing theft, the HR treatment does not significantly reduce stealing relative to the NV baseline. This result implies that simply prompting individual moral reflection is insufficient to shift behavior towards the norm proscribing theft or altruistic motives more generally, at least under substantial normative ambiguity. The RR treatment, by contrast, turns out to be relatively effective at mitigating theft compared to the NV and HR treatments. This finding underscores the critical role of publishing inter-individual support for the course of action considered appropriate in a specific context. Self-nudging or subtle interventions designed as unincentivized commitment devices that have been shown to be effective in some contexts (see Reijula & Hertwig, 2022; Tontrup & Sprigman, 2022) seem to be rather ineffective at curbing selfish behavior.

Our analysis of heterogeneous treatment effects uncovers part of the behavioral pattern engendered by our voting procedures. The RR treatment only entails a significant reduction in theft relative to the NV treatment at the lower quantile (20th percentile) of the stealing distribution, while participants at higher quantiles remain unaffected. Assuming that participants

at the lower quantile incur higher moral costs of theft than participants at higher quantiles, it seems that publicizing the outcome only shifts the behavior of people whose costs of breaking the social norm prohibiting theft are high. This suggests that the democracy premium particularly resonates with individuals already predisposed to comply with injunctive norms.

The reduction in theft we observe in the RR treatment raises the question whether this effect depends on the strength of the democratic support expressed for the social norm proscribing theft or the social norm of merit. To our surprise, the aggregate strength of democratic support for prohibiting theft does not explain variation in stealing decisions. One plausible interpretation is that participants draw on pre-existing social norms to interpret the ambiguous rules, rather than relying solely on the observed vote.[6]

Finally, an exploratory analysis of the HR treatment suggests that voting in private may push people to exploit moral wiggle room. Participants who explicitly reject the norm of merit (i.e., favor the prohibition on theft) often proceed to steal large amounts nonetheless. This discrepancy between expressed support for the norm proscribing theft and behavior hints at a form of moral licensing in which casting an altruistic vote privately allows individuals to preserve a positive self-image while engaging in violations of norms prohibiting selfish behavior.

Our study does not come without limitations. While we do contribute to the literature on the democracy premium and expressive law, our experiment is not designed to generate findings transposable to any specific democratic system. We therefore refrain from making general claims about the effect of publicity in specific institutional contexts or in democracies writ large.

Implications. Our study highlights two critical points. First, democratic engagement alone – absent public acknowledgment of the voting outcome – fails to resolve normative ambiguity or encourage altruistic behavior. Second, making collective decisions visible appears essential for leveraging the democracy premium. This has direct implications for policymakers and legislators seeking to leverage expressive law: if the goal is to reduce norm violations (e.g., theft), then ensuring high visibility of democratic support for the norm proscribing theft is crucial.

At the same time our study indicates that purely introspective or self-nudging procedures may risk backfiring in contexts characterized by normative ambiguity. When moral wiggle room can be exploited due to a certain degree of secrecy, participants can maintain a positive self-image while still engaging in selfish acts. Publicizing democratic outcomes reduces this wiggle room, motivating individuals – especially those who incur relatively high costs of breaking the law – to follow the law or norm in question.

The expressive power of law rests on challenging behavioral requirements. Under norma-

[6] An analysis of responses to our post-experimental survey, particularly from participants who rejected the social norm of merit and took either 8, 9, or 10 points, reveals a clear cognitive dissonance: while many of these participants propose arguments in support of the prohibition of theft, they do steal large parts of the victims' endowment.

tive ambiguity, democratic votes can indeed bolster altruistic behavior, but only when the collective decision is made salient. Concealed voting outcomes can inadvertently invite hypocrisy and moral licensing, as individuals may endorse one norm just to break it the next moment. Institutions and legislators thus face a key design challenge: if they seek to harness the expressive power of democratic procedures, publicizing the collective will is key. A *res publica* supporting obedience with the law truly earns its name only when the outcome of the legislative process is sufficiently public.

References

Baldassarri, D., & Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11023–11027.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, *76*(2), 169–217.

Bicchieri, C. (2005). *The grammar of society*. Cambridge University Press.

Bosco, L. (2022). The moral wiggle room: When avoiding information is strategically optimal. *International Journal of Buisness and Social Science*, *13*(1), 20–27.

Casoria, F., Galeotti, F., & Villeval, M. C. (2021). Perceived social norm and behavior quickly adjusted to legal changes during the covid-19 pandemic. *Journal of Economic Behavior & Organization*, *190*, 54–65.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (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*(6), 1015–1026.

Cooter, R. (1998). Expressive law and economics. *The Journal of Legal Studies*, *27*(S2), 585–607.

Dal Bó, P. (2014). Experimental evidence on the workings of democratic institutions. In S. Galiani & I. Sened (Eds.), *Institutions, property rights, and economic growth* (pp. 266–288). Cambridge University Press.

Dal Bó, P., Foster, A., & Kamei, K. (2019). The democracy effect: A weights-based identification strategy.

Dal Bó, P., Foster, A., & Putterman, L. (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *American Economic Review*, *100*(5), 2205–2229.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, *100*(2), 193–201.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.

Depoorter, B., & Tontrup, S. (2012). How law frames moral intuitions: The expressive effect of specific performance. *Arizona Law Review*, *54*, 673.

Dharmapala, D., & McAdams, R. H. (2003). The condorcet jury theorem and the expressive function of law: A theory of informative law. *American Law and Economics Review*, *5*(1), 1–31.

Engel, C. (2016). A random shock is not random assignment. *Economics Letters*, *145*, 45–47.

Engel, C., Heine, K., & Naseer, S. (2021). Religion and tradition in conflict: Experimentally testing the power of social norms to invalidate religious law.

Engel, C., & Nagin, D. (2015). Who is afraid of the stick? experimentally testing the deterrent effect of sanction certainty. *Review of Behavioral Economics*, *2*(4), 405–434.

Faillo, M., Rizzolli, M., & Tontrup, S. (2019). Thou shalt not steal: Taking aversion with legal property claims. *Journal of Economic Psychology*, *71*, 88–101.

Feldman, Y. (2006). The expressive function of the law: Legality, cost, intrinsic motivation and consensus.

Feldman, Y., & Nadler, J. (2006). Expressive law and file sharing norms. *San Diego Law Review*, *43*, 577–618.

Flage, A. (2024). Taking games: A meta-analysis. *Journal of the Economic Science Association*, 1–24.

Funk, P. (2007). Is there an expressive function of law? an empirical analysis of voting laws with symbolic fines. *American Law and Economics Review*, *9*(1), 135–159.

Galbiati, R., Henry, E., Jacquemet, N., & Lobeck, M. (2021). How laws affect the perception of norms: Empirical evidence from the lockdown. *PLoS One*, *16*(9), e0256624.

Galbiati, R., & Vertova, P. (2014). How laws affect behavior: Obligations, incentives and cooperative behavior. *International Review of Law and Economics*, *38*, 48–57.

Görges, L., Lane, T., Nosenzo, D., & Sonderegger, S. (2023). Equal before the (expressive power of) law?

Gravert, C. (2013). How luck and performance affect stealing. *Journal of Economic Behavior & Organization*, *93*, 301–304.

Grossman, G., & Baldassarri, D. (2012). The impact of elections on cooperation: Evidence from a lab-in-the-field experiment in uganda. *American journal of political science*, *56*(4), 964–985.

Henkin, L. (1979). *How nations behave: Law and foreign policy* (2. ed.). Columbia University Press.

Henrich, J., Boyd, R., Bowles, S., Camerer, C. F., Fehr, E., Gintis, H., & McElreath, R. (2004). Overview and synthesis. In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, & H. Gintis (Eds.), *Foundations of human sociality* (pp. 8–54). Oxford University Press.

Hermstrüwer, Y., & Dickert, S. (2017). Sharing is daring: An experiment on consent, chilling effects and a salient privacy nudge. *International Review of Law and Economics*, *51*, 38–49.

Kamei, K. (2016). Democracy and resilient pro-social behavioral change: An experimental study. *Social Choice and Welfare*, *47*(2), 359–378.

Khadjavi, M., & Lange, A. (2015). Doing good or doing harm: Experimental evidence on giving and taking in public good games. *Experimental Economics*, *18*, 432–441.

Korenok, O., Millner, E. L., & Razzolini, L. (2014). Taking, giving, and impure altruism in dictator games. *Experimental Economics*, *17*, 488–500.

Korenok, O., Millner, E. L., & Razzolini, L. (2018). Taking aversion. *Journal of Economic Behavior & Organization*, *150*, 397–403.

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

Lane, T., Nosenzo, D., & Sonderegger, S. (2023). Law and norms: Empirical evidence. *American Economic Review*, *113*(5), 1255–1293.

Langenbach, P., & Tausch, F. (2019). Inherited institutions: Cooperation in the light of democratic legitimacy. *Journal of Law, Economics, and Organization*, *35*(2), 364–393.

Larson, T., & Capra, C. M. (2009). Exploiting moral wiggle room: Illusory preference for fairness? a comment. *Judgment and Decision Making*, *4*(6), 467–474.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21*(2), 153–174.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, *115*(3), 482–493.

Marcin, I., Robalo, P., & Tausch, F. (2019). Institutional endogeneity and third-party punishment in social dilemmas. *Journal of Economic Behavior & Organization*, *161*, 243–264.

McAdams, R. H. (2000a). An attitudinal theory of expressive law. *Oregon Law Review*, *79*, 339.

McAdams, R. H. (2000b). A focal point theory of expressive law. *Virginia Law Review*, *86*(8), 1649.

McAdams, R. H. (2015). *The expressive powers of law: Theories and limits*. Harvard University Press.

McAdams, R. H., & Nadler, J. (2005). Testing the focal point theory of legal compliance: The effect of third–party expression in an experimental hawk/dove game. *Journal of Empirical Legal Studies*, *2*(1), 87–123.

McAdams, R. H., & Nadler, J. (2008). Coordinating in the shadow of the law: Two contextualized tests of the focal point theory of legal compliance. *Law & Society Review*, *42*(4), 865–898.

McAdams, R. H., & Rasmusen, E. B. (2007). Chapter 20: Norms and the law. In S. Shavell & A. M. Polinsky (Eds.), *Handbook of law and economics* (pp. 1573–1618, Vol. 2). Elsevier.

Nadler, J. (2017). Expressive law, social norms, and social groups. *Law & Social Inquiry*, *42*(1), 60–75.

Nadler, J. (2024). Expressive law and social norms. In R. Hollander-Blumoff (Ed.), *Research handbook on law and psychology* (pp. 328–342). Edward Elgar Publishing.

Posner, E. A. (1998). Symbols, signals, and social norms in politics and the law. *The Journal of Legal Studies*, *27*(S2), 765–797.

Posner, E. A. (2002). The signaling model of social norms: Further thoughts. *University of Richmond Law Review*, *36*, 465.

Rawls, J. (1971). *A theory of justice*. Belknap Press.

Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. *Behavioural Public Policy*, *6*(1), 119–149.

Rizzolli, M., & Stanca, L. (2012). Judicial errors and crime deterrence: Theory and experimental evidence. *Journal of Law and Economics*, *55*(2), 311–338.

Schauer, F. F. (2015). *The force of law*. Harvard University Press.

Stok, F. M., & de Ridder, D. T. D. (2019). The focus theory of normative conduct. In K. Sassenberg & M. L. W. Vliek (Eds.), *Social psychology in action* (pp. 95–110). Springer International Publishing.

Sunstein, C. R. (1996). On the expressive function of law. *University of Pennsylvania Law Review*, *144*(5), 2021–2053.

Sutter, M., Haigner, S., & Kocher, M. G. (2010). Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations. *Review of Economic Studies*, *77*(4), 1540–1566.

Teichman, D., & Leibovitch, A. (2024). Normative ambiguity, social norms and the expressive power of law. *Unpublished Manuscript*.

Thunström, L., Cherry, T. L., McEvoy, D. M., & Shogren, J. F. (2016). Endogenous context in a dictator game. *Journal of Behavioral and Experimental Economics*, *65*, 117–120.

Tontrup, S., & Gaissmaier, W. (2023). Democracy premium: A cross-cultural comparison of effects of democratic choice in china and germany. *SSRN Electronic Journal*.

Tontrup, S., & Sprigman, C. J. (2022). Self-nudging contracts and the positive effects of autonomy—analyzing the prospect of behavioral self-management. *Journal of Empirical Legal Studies*, *19*(3), 594–676.

Tyler, T. R. (2006). *Why people obey the law*. Princeton University Press.

Tyler, T. R. (2010). *Why people cooperate: The role of social motivations*. Princeton University Press.

Tyran, J.-.-R., & Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent. *The Scandinavian Journal of Economics*, *108*(1), 135–156.

van der Weele, J. J., Kulisa, J., Kosfeld, M., & Friebel, G. (2014). Resisting moral wiggle room: How robust is reciprocal behavior? *American Economic Journal: Microeconomics*, *6*(3), 256–264.

A Appendix

A.1 Treatment Effects

Amount Stolen. Tables 6 and 7 show the summary statistics of the amount stolen across all treatments and the results of a two-sided Mann-Whitney U test respectively.

Table 6: Summary statistics of amount stolen

Treatment	**No. Obs.**	**Mean**	**SD**	**Min**	**25%**	**Median**	**75%**	**Max**
Total	240	5.33	3.83	0.00	2.00	5.00	10.00	10.00
No Voting	60	5.78	3.69	0.00	3.00	5.00	10.00	10.00
Hidden Result	60	6.18	3.64	0.00	4.00	6.00	10.00	10.00
Revealed Result	120	4.67	3.90	0.00	0.75	4.50	9.00	10.00

Table 7: Results of a Two-Sided Mann-Whitney U test

Comparison	U-Value	p
No Voting vs. Hidden Result	1669.5	0.485
No Voting vs. Revealed Result	4190.5	0.068
Hidden Result vs. Revealed Result	4402.0	0.013

Fig. 5 plots the coefficients of Model 3 in Tables 1 and 2 respectively.

Figure 5: Coefficient plots for Model 3 in Table 1 and 2

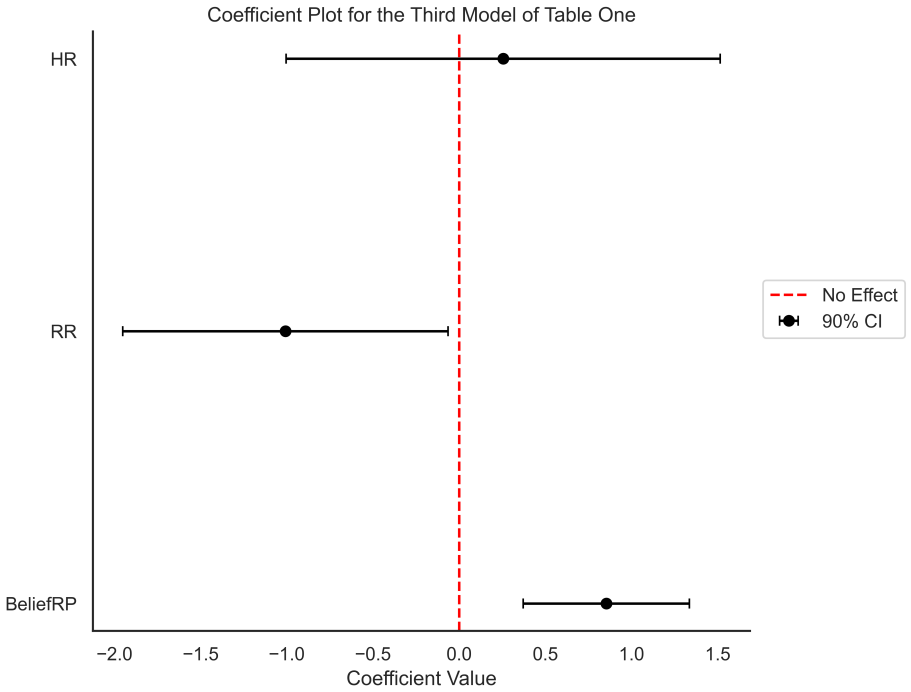


Table 1

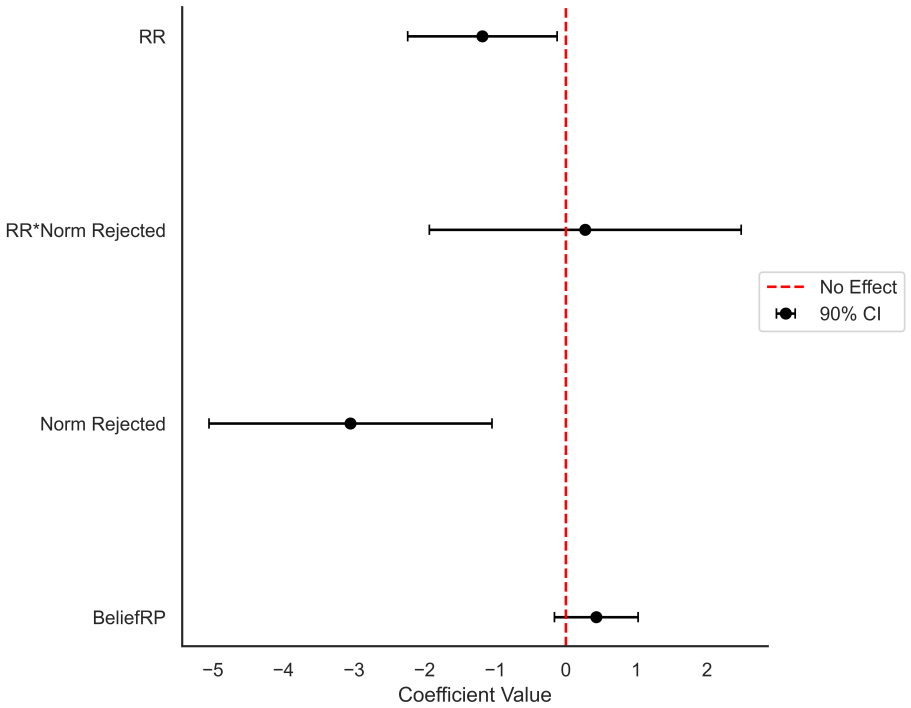


Table 2

Table 8 shows the results of a linear mixed model.

Table 8: Treatment effects

DV: Amount stolen	(1)	(2)	(3)
Revealed Result	-1.117*	-1.069*	-1.008*
	(0.598)	(0.601)	(0.593)
Hidden Result	0.400	0.388	0.255
	(0.691)	(0.686)	(0.678)
Performance Belief			0.854***
			(0.313)
Sociodemographics		✓	✓
Intercept	5.783***	6.898***	6.194***
	(0.489)	(1.499)	(1.500)
Observations	240	240	240
Residual Std. Error	3.784 (df=237)	3.688 (df=227)	3.636 (df=226)

This table reports the results of a linear mixed model. Standard errors are reported in parentheses. The *No Voting* treatment serves as the reference category.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Mediation Analysis. Table 9 shows the results of a mediation analysis.

Table 9: Mediation analysis

Path	**Coefficient**	**Standard Error**	**p-value**	**CI [2.5%]**	**CI [97.5%]**	**Significant**
Performance Belief → Norm Rejection [*]	-0.177	0.051	0.000	-0.277	-0.078	Yes
Norm Rejection → Amount Stolen [**]	-2.919	0.554	0.000	-4.004	-1.833	Yes
Total	0.966	0.356	0.007	0.269	1.664	Yes
Direct	0.507	0.342	0.138	-0.164	1.178	No
Indirect [***]	0.518	0.100	0.000	0.318	0.708	Yes

[*] Effect of independent variable on mediator.
[**] Effect of mediator on dependent variable controlling for independent variable.
[***] Indirect effects are reported with bootstrapped confidence intervals.

Proportion of Theft. Fig. 6 plots the proportion of theft across all treatments. Table 10 shows the results of a logistic regression model.

Figure 6: Proportion of theft across treatments

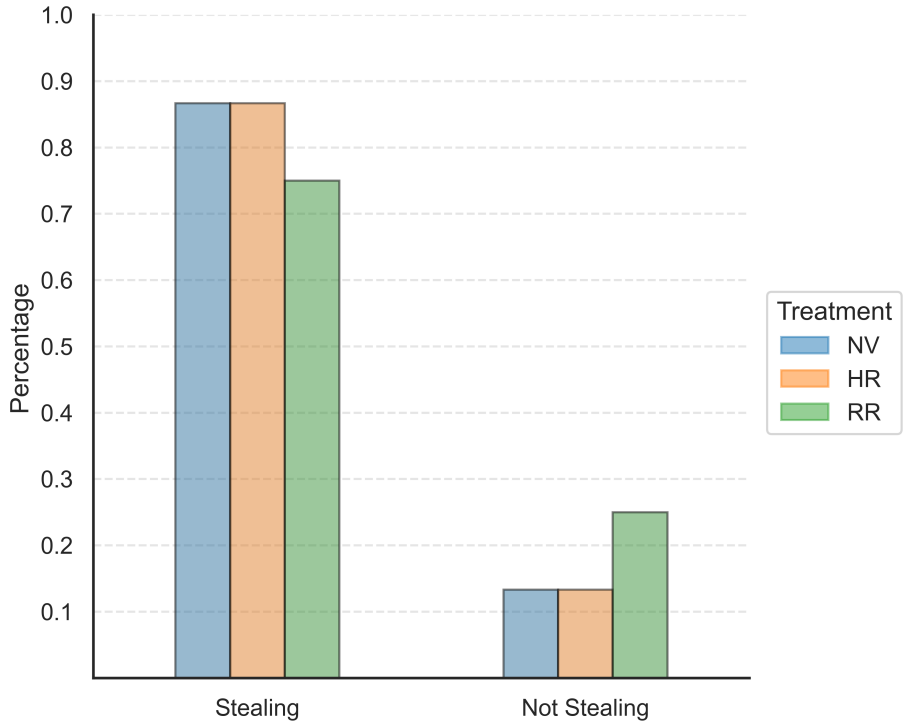


Table 10: Treatment effects

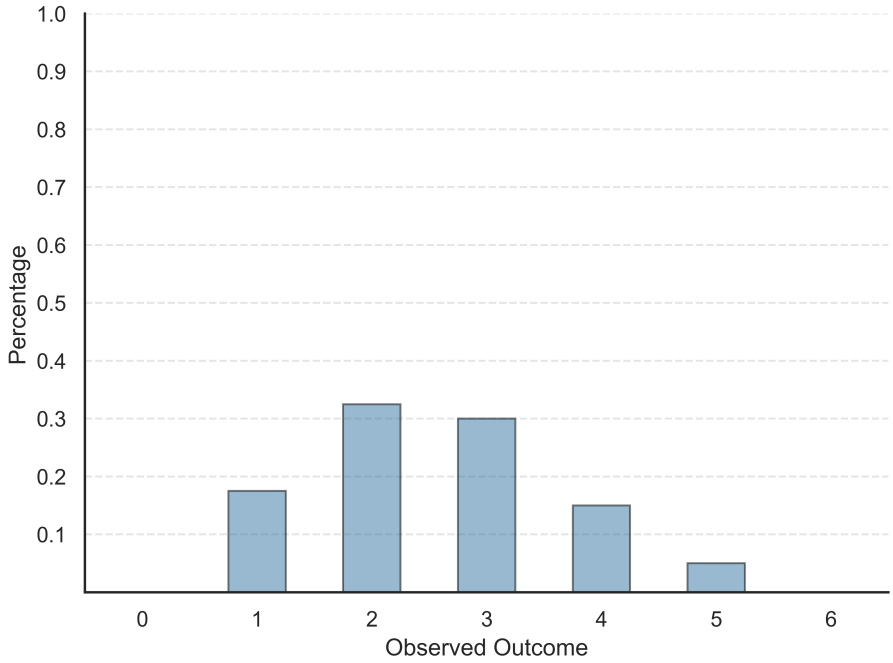
DV: Stealing Dummy	(1)	(2)	(3)	(1)	(2)	(3)
Hidden Result	-0.000	0.022	0.005			
	(0.611)	(0.683)	(0.689)			
Revealed Result	-0.116*	-0.107*	-0.103*	-0.119	-0.124	-0.109
	(0.438)	(0.444)	(0.439)	(0.633)	(0.985)	(0.967)
Revealed Result x Norm Rejected					0.045	0.027
					(1.195)	(1.202)
Norm Rejected					-0.244	-0.066
					(1.169)	(1.311)
Performance Belief			0.068**	0.060*	0.027	0.136**
			(0.218)	(0.237)	(0.269)	(0.468)
Performance Belief x Norm Rejected						-0.158**
						(0.597)
Sociodemographics	✓	✓	✓	✓	✓	✓
Constant	1.872***	1.481	1.016	1.316	2.691*	1.921
	(0.380)	(1.125)	(1.203)	(1.424)	(1.524)	(1.533)
Observations	240	240	240	180	180	180
Pseudo R^2	0.023	0.086	0.105	0.110	0.179	0.199

This table reports the results of a logistic regression model. Standard errors in parentheses and clustered at the group level. All coefficients are reported as marginal effects. The *No Voting* treatment serves as the reference category in all model specifications.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Votes. Fig. 7 plots the distribution of votes considering theft to be socially appropriate.

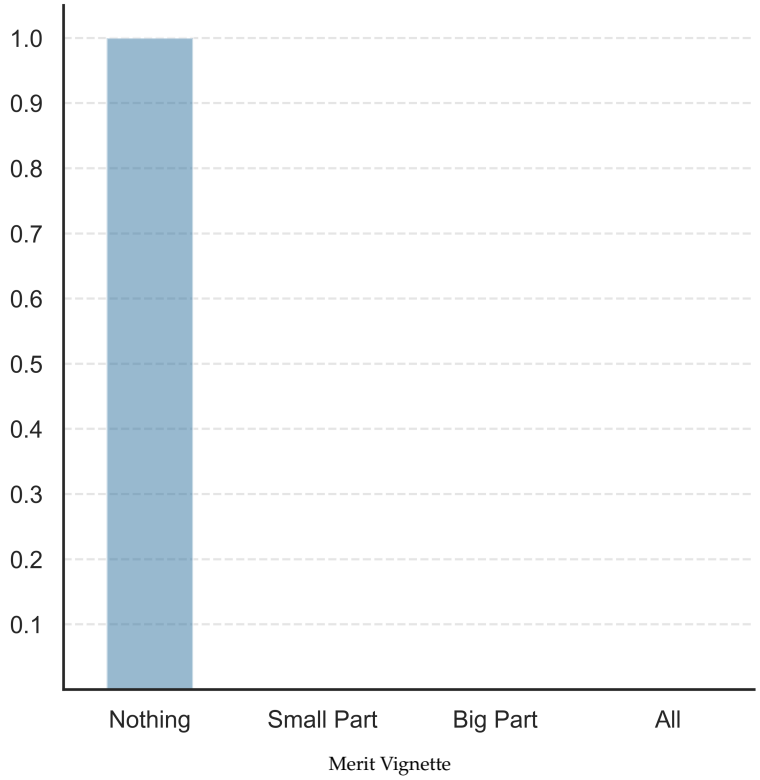
Figure 7: Observed outcome



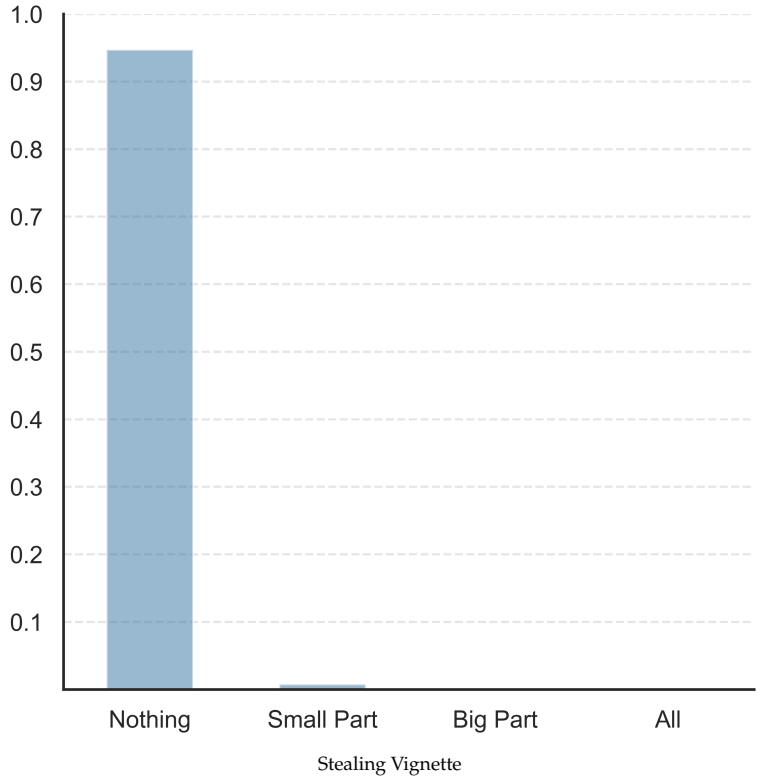
A.2 Norm Elicitation Experiment

Fig. 8 shows that abstaining from theft is considered socially appropriate both when taking can be justified based on merit and when no such justification is available.

Figure 8: Percentage of participants describing theft as socially appropriate



Merit Vignette



Stealing Vignette

A.3 Reasoning Process

In all our treatments, participants faced a moral dilemma arising from the conflict between the social norm proscribing theft and the social norm of merit justifying theft. To understand the reasoning process adopted by participants in addressing this dilemma, we implemented an open-form questionnaire asking for the justifications brought forward by participants in support of their behavior in the stealing game and of their vote. Our descriptive analysis of responses shows that participants were well aware of the normative ambiguity they faced in the experiment and highlights a broad variety of arguments explaining participants' stealing and voting behavior.

Justification for Stealing.

- No Voting Treatment
  - Favor
    - Self-Interest and Financial Need
    - Perceived Merit or Effort
    - Maximizing Gain
  - Against
    - Moral Discomfort
    - Fairness and Equality
    - Empathy
- Hidden Result Treatment
  - Favor
    - Maximizing Personal Gain
    - Perceived Merit or Effort
    - Anonymity and Lack of Consequences
    - Perceived Fairness Based on Winning in Competition
  - Against
    - Moral Discomfort
    - Fairness and Equality
    - Empathy and Consideration for Others
    - Reluctance to Abuse Power
- Revealed Result Treatment
  - Favor
    - Maximizing Personal Gain
    - Perceived Merit or Effort
    - Opportunism
    - Norms and Expectations Elicited from the Design
    - Financial Need
  - Against

        * Fairness and Equality
        * Moral Discomfort
        * Guilt and Empathy
        * Reluctance to Abuse Power
        * Randomness and Luck

Justification for Votes.

- Hidden Result Treatment
    - Approve
        * Performance-based Justification
        * Self-Interest and Strategy
        * Rule Following (Status quo)
    - Reject
        * Fairness and Equity
        * Moral and Ethical Considerations
        * Uncertainty or Risk Mitigation
- Revealed Result Treatment
    - Approve
        * Performance-Based Justification
        * Self-Interest and Strategy
        * Curiosity and Excitement
    - Reject
        * Fairness and Equity
        * Moral and Ethical Considerations
        * Uncertainty or Risk Mitigation

A.4 Instructions

INTRODUCTION [ALL TREATMENTS]

Please give this study your full attention. You will have a limited amount of time to complete it.

You can earn money in this study. Your earnings depend on your decisions and those made by other participants. The sum of your earnings in this part of the experiment and in the previous part that was held on (write the date) will be transferred to your bank account within the next few days following the study. In addition, you will receive a flat fee of 2.5 Euros for showing up on time. If you leave the study before it ends, you will not receive any payment.

Your earnings are given in points. At the end of the study, you will be paid based on the following exchange rate:

1 Point = 0.6 Euros

In this study, you are randomly assigned to a group of 6 participants and randomly paired with another participant within your group. You and the participant you are paired with will be assigned to one of two roles: Person A or Person B. The study, however, is completely anonymous. The other participants will not be informed about your name, your decisions, or your payment.

If you have any questions during the study, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. Participants intentionally violating the rules may be asked to leave the experiment and may not be paid.

Note: As you can see on top of this screen, these instructions are organized in different tabs (Introduction, Main Decision, Procedure, Control Questions). You can switch back and forth between these tabs. All tabs except the tab with the Practice Questions will be accessible during the entire session.

MAIN DECISION [ALL TREATMENTS]

Participants will be endowed with 10 points, regardless of whether they have been assigned the role of Person A or of Person B.

Person A will decide how many points they want to take from the endowment of Person B. Person A can choose any amount between 0 and 10 points. Person A receives the sum of their own endowment and the amount taken from Person B. Person B receives their endowment

minus the number of points that Person A has taken. For example, if Person A takes 4 points from Person B, Person A receives 14 points (i.e., 10 points of their own endowment and 4 points taken from the endowment of Person B), and Person B receives 6 points.

PROCEDURE [BASELINE]

The experiment proceeds in four stages.

Stage 1 You and the participant you are paired with will compete in a task. The task consists of counting the number of zeros in tables of 150 randomly ordered zeros and ones. The figure shows the work screen you will use later:

[SCREEN]

Enter the number of zeros into the box on the right side of the screen. After you have entered the number, click the OK button. If you enter the correct result, a new table will be generated. You can see the total number of tables correctly solved while you are doing the task.

If your input is wrong, you have two additional tries to enter the correct number into the table. You therefore have a total of three tries to solve each table. If you enter three times a wrong number, a new table will then be generated.

You have 5 minutes for this task. The remaining time is displayed in the upper right-hand corner of the screen.

The participant solving more tables correctly within the 5 minutes available for the task will be assigned the role of Person A. The other participant will be assigned the role of Person B. If you and the participant you are paired with managed to solve the same number of problems, roles will be assigned randomly.

Stage 2 Then, if you have been assigned the role of Person A, you will be asked about your relative performance compared to the participant you are paired with. If you believe that you managed to solve more problems than the participant you are paired with, you can indicate your belief about the difference between your performance and the performance of the other participant with by typing a number (e.g., 3 if you believe you managed to solve 3 more problems than the other participant). If you believe that you and the other participant solved the same number of problems and roles were assigned randomly, you can indicate your belief by typing 0.

Stage 3 In this stage, if you have been assigned the role of Person A, you will make a decision according to the rules described in the "Main Decision" tab.

Stage 4 You will fill out a questionnaire at this stage and by the submitting it the study ends.

PROCEDURE [HIDDEN RESULT TREATMENT]

The experiment proceeds in five stages.

Stage 1 You and the participant you are paired with will compete in a task. The task consists of counting the number of zeros in tables of 150 randomly ordered zeros and ones. The figure shows the work screen you will use later:

[SCREEN]

Enter the number of zeros into the box on the right side of the screen. After you have entered the number, click the OK button. If you enter the correct result, a new table will be generated. You can see the total number of tables correctly solved, while you are doing the task.

If your input is wrong, you have two additional tries to enter the correct number into the table. You therefore have a total of three tries to solve each table. If you enter three times a wrong number, a new table will then be generated.

You have 5 minutes for this task. The remaining time is displayed in the upper right-hand corner of the screen.

The participant solving more tables correctly within the 5 minutes available for the task will be assigned the role of Person A. The other participant will be assigned the role of Person B. If you and the participant you are paired with managed to solve the same number of problems, roles will be assigned randomly. However, you will only learn about your role in Stage 3.

Stage 2 In this stage, you will be asked to vote on the following norm:

"Person A is allowed to take what they deserve from Person B."

You will be able to decide in favour of this norm by clicking "YES" or against this norm by clicking "NO".

Stage 3 At the beginning of this stage, you will be informed about the role you have been assigned to in Stage 1.

Then, if you have been assigned the role of Person A, you will be asked about your relative performance compared to the participant you are paired with. If you believe that you managed to solve more problems than the participant you are paired with, you can indicate your belief about the difference between your performance and the performance of the other participant with by typing a number (e.g., 3 if you believe you managed to solve 3 more problems than the other participant). If you believe that you and the other participant solved the same number of problems and roles were assigned randomly, you can indicate your belief by typing 0.

In addition, if you have been assigned the role of Person A, you will be asked to indicate your belief about the number of votes in favor of the norm in your group. For example, if you believe that 1 person voted in favor of the norm, enter 1 in the designated input box.

Stage 4 In this stage, if you have been assigned the role of Person A, you will make a decision according to the rules described in the "Main Decision" tab.

Stage 5 You will fill out a questionnaire at this stage and by the submitting it the study ends.

Before you begin with the actual experiment, it is important for you to know how the experiment proceeds. We would therefore ask you please to answer some control questions on the experiment. As soon as all participants have correctly answered these questions, the experiment will begin.

PROCEDURE [REVEALED RESULT TREATMENT]

The experiment proceeds in five stages.

Stage 1 You and the participant you are paired with will compete in a task. The task consists of counting the number of zeros in tables of 150 randomly ordered zeros and ones. The figure shows the work screen you will use later:

[SCREEN]

Enter the number of zeros into the box on the right side of the screen. After you have entered the number, click the OK button. If you enter the correct result, a new table will be generated. You can see the total number of tables correctly solved, while you are doing the task.

If your input is wrong, you have two additional tries to enter the correct number into the table. You therefore have a total of three tries to solve each table. If you enter three times a wrong number, a new table will then be generated.

You have 5 minutes for this task. The remaining time is displayed in the upper right-hand corner of the screen.

The participant solving more tables correctly within the 5 minutes available for the task will be assigned the role of Person A. The other participant will be assigned the role of Person B. If you and the participant you are paired with managed to solve the same number of problems, roles will be assigned randomly. However, you will only learn about your role in Stage 3.

Stage 2 In this stage, you will be asked to vote on the following norm:

"Person A is allowed to take what they deserve from Person B."

You will be able to decide in favour of this norm by clicking "YES" or against this norm by clicking "NO".

You will see on your screen how many of your group members vote in favour of the norm, after all your group members cast their votes.

Stage 3 At the beginning of this stage, you will be informed about the role you have been assigned to in Stage 1.

Then, if you have been assigned the role of Person A, you will be asked about your relative performance compared to the participant you are paired with. If you believe that you managed to solve more problems than the participant you are paired with, you can indicate your belief about the difference between your performance and the performance of the other participant with by typing a number (e.g., 3 if you believe you managed to solve 3 more problems than the other participant). If you believe that you and the other participant solved the same number of problems and roles were assigned randomly, you can indicate your belief by typing 0.

Stage 4 In this stage, if you have been assigned the role of Person A, you will make a decision according to the rules described in the "Main Decision" tab.

Stage 5 You will fill out a questionnaire at this stage and by submitting it the study ends.