FROM JUDGMENT TO ACTION:

TOWARD A BEHAVIORAL PERSPECTIVE ON LEGITIMACY

Moritz Gruban
Royal Holloway, University of London
moritz.gruban@rhul.ac.uk

Matthieu Légeret
University of Amsterdam
m.legeret@uva.nl

University of Lausanne
matthieu.legeret@unil.ch

Patrick Haack
(Corresponding author)
University of Lausanne
patrick.haack@unil.ch

March 3, 2026

Paper forthcoming in the ***Journal of Management***:
https://doi.org/10.1177/01492063261416429

Supplemental material for this article is available at http://xxx.sagepub.com/supplemental.

Acknowledgments

We thank our editor, Peggy Lee, for her support and constructive guidance. We also thank the two anonymous reviewers and the JOM method expert for their valuable feedback throughout the review process. We are also grateful to Dianna Amasino, Alex Bitektine, Neil Conway, Aurélien Feix, Derek Harmon, Jan Hausfeld, Anna Jasinenko, Sirio Lonati, Maya Modhir, Déborah Philippe, Thomas Roulet, Oliver Schilke, Jonas Spengler, Roy Suddaby, Leigh Tost, Georg Wernicke, and Christian Zehnder for their helpful comments on earlier versions of the manuscript. Moritz Gruban and Matthieu Légeret gratefully acknowledge the funding they received from the Swiss National Science Foundation (grant number 206548 and grant number 210827).

Abstract

Research grounded in the legitimacy-as-perception perspective has primarily conceptualized legitimacy as a socio-cognitive judgment, assuming that evaluators act in line with their propriety judgments by supporting legitimacy objects they deem appropriate and opposing those they judge as inappropriate. However, this assumed alignment between legitimacy judgments and behaviors has rarely been tested directly. Advancing a behavioral perspective of legitimacy therefore requires systematically examining when and why legitimacy manifests in action rather than judgment alone. We advance this agenda through two experiments that disentangle evaluative and behavioral responses to legitimacy objects. Study 1 (MTurk; n = 1,889) manipulates self-interest through incentives, norm congruence through framing, and whether propriety judgments are measured prior to behavioral responses or not. Study 2 (Prolific; n = 801) replicates this design, focusing on norm congruence and propriety judgment measurement prior to behavioral responses. Our results reveal several inconsistencies between the evaluative and behavioral manifestations of legitimacy. Specifically, the effects of norm violations and self-interest on behavior diverge from their effects on propriety judgments. In addition, Study 1 reveals a "neutralization effect," whereby eliciting propriety judgments attenuates behavioral responses, an effect that was not observed in Study 2. Together, these findings challenge the assumption that legitimacy judgments reliably translate into behavior and underscore the need to distinguish between evaluative and behavioral manifestations of legitimacy. By identifying conditions under which legitimacy judgments guide, or fail to guide, action, this study advances a behavioral perspective on legitimacy and offers methodological guidance for future experimental research.

Keywords: legitimacy; propriety judgments; behavior; norms; self-interest; experimental research

INTRODUCTION

Legitimacy, defined as a judgment "of the extent to which an entity is appropriate for its social context" (Tost, 2011: 689), is a key concept in organization and management studies (Greenwood et al., 2017). A burgeoning perspective conceptualizes legitimacy as a socio-cognitive process, focusing on how evaluators form and express their judgments (Bitektine, 2011; Bitektine & Haack, 2015; Haack, Schilke, & Zucker, 2021). Within this "legitimacy-as-perception" perspective (Suddaby, Bitektine, & Haack, 2017), researchers posit that legitimacy manifests in judgments and behaviors that tend to be consistent. Specifically, when evaluators judge a legitimacy object, such as an organization, a practice, or an action, as appropriate, they support it. Conversely, evaluators sanction or seek to change objects that they deem as inappropriate (Bitektine, 2011; Tost, 2011). The Volkswagen emissions scandal exemplifies a consistent pattern between judgments and behaviors. Not only did regulators, consumer protection agencies, and customers widely condemn Volkswagen's fraudulent actions, but the company also had to pay more than $30 billion in fines and damages, and its sales decreased in key markets (Jolly, 2019).

However, alignment between judgments and behavior is not guaranteed. In fact, research has shown that judgments and intentions often deviate from actual behaviors (Ajzen, Brown, & Carvajal, 2004; Sheeran & Webb, 2016), especially in morally charged contexts (Effron & Helgason, 2023). For example, consumers often criticize fast fashion companies for their negative environmental impact and poor labor practices, yet brands such as Primark and Shein continue to thrive (McKinsey, 2025). In this case, moral evaluations and market behaviors diverge, revealing a gap between legitimacy judgments and behavioral support.

Together, these examples illustrate that, although legitimacy judgments often influence behavior, this relationship is neither uniform nor predictable. Sometimes, evaluators act in line with their judgments; at other times, however, moral criticism coexists with continued support.

These contrasting patterns raise a fundamental question: when and why do legitimacy judgments correspond with behavior, and when do they not? Studying whether and when such correspondence occurs is crucial for understanding how legitimacy shapes organizational and institutional outcomes (Deephouse et al., 2017; Suddaby et al., 2017; Tost, 2011). More broadly, addressing this question is essential for developing a behavioral perspective of legitimacy, which moves beyond legitimacy as an evaluative assessment to examine how and when legitimacy judgments are translated into concrete actions, such as support, compliance, sanctioning, or withdrawal.

To address the question of when and why legitimacy judgments align, or fail to align, with behavior, we conducted two experimental studies. The first study was a large-scale experiment with 1,889 participants recruited via Amazon Mechanical Turk (MTurk) that followed best-practice recommendations from the literature (Aguinis, Villamor, & Ramani, 2021). The second study was a preregistered replication on Prolific with 801 participants, designed to assess the robustness and generalizability of the findings of the first study across independent samples.

Building on the legitimacy-as-perception perspective, we conceptualized decisions about the allocation of resources between personal and prosocial interests as the focal legitimacy object. In both studies, we examined how individual evaluators, such as consumers or citizens, form propriety judgments and respond behaviorally to allocation decisions, thereby addressing calls to investigate the behavioral manifestations of legitimacy (Bitektine et al., 2020; Haack et al., 2021). We manipulated two core antecedents of legitimacy: self-interest and norm congruence (Ashforth & Gibbs, 1990; Bitektine, 2011; Suchman, 1995; Tost, 2011) and assessed their effects on both evaluative and behavioral manifestations of legitimacy. The results of both studies show that judgments and behaviors are not always aligned. Furthermore, we tested whether expressing a propriety judgment influences subsequent behavior by

comparing conditions in which judgments were or were not elicited prior to behavioral responses. This comparison revealed that, on MTurk, expressing a judgment appeared to *neutralize* behavioral reactions, that is, it absorbed the effects of self-interest and norm congruence on subsequent behavioral support. However, we did not observe this pattern on Prolific. We discuss these results in light of real-world examples that illustrate consistent and inconsistent legitimacy judgments and actions.

Our contributions are threefold. First, we advance a behavioral perspective by shifting the focus of legitimacy research on action and demonstrating how behavioral manifestations can complement insights on evaluative judgments. In our MTurk study, evaluators often act toward the legitimacy object when we measure only behavioral responses, most notably by punishing decisions framed as norm-violating. The replication study on Prolific corroborated these framing effects on behavior. At the same time, we found that actions were not always consistent with propriety judgments, which emphasizes the importance of distinguishing conceptually and methodologically between the evaluative and behavioral manifestations of legitimacy. Second, our comparison of conditions with and without prior measurement of propriety judgments on MTurk suggests that expressing a judgment may suppress subsequent behavioral responses. We term this phenomenon the "neutralization effect." This effect did not emerge on Prolific, suggesting that it may be contingent on the evaluating audience or on contextual factors shaping how legitimacy judgments are formed and expressed. Third, we make a methodological contribution by responding to calls for more mechanism-based experimental studies in legitimacy research (Deephouse et al., 2017; Haack et al., 2021; Suddaby et al., 2017) and organization theory more broadly (Levine et al., 2023). Our design simulates real-world decision-making contexts and exogenously manipulates whether participants engage in behavior only or both judgment and behavior, addressing endogeneity concerns. While prioritizing internal validity to identify behavioral mechanisms, we enhance

external relevance by using realistic decision scenarios that evaluators, such as consumers and citizens, face. Together, these choices provide a blueprint for future behavioral studies of legitimacy and underscore the need to refine both theory and methodology (Bergh et al., 2022).

THEORETICAL BACKGROUND

The study of legitimacy is a foundational area of management and organization studies, particularly within institutional theory (Deephouse et al., 2017; Greenwood et al., 2017; Suddaby et al., 2017). Traditionally, this body of work has focused on legitimacy at the collective level, maintaining that legitimacy emerges as an irreducible whole from social interaction and is thus largely "independent of particular observers" (Suchman, 1995: 574). More recently, scholars have complemented this collective-level understanding of legitimacy by theorizing how individuals form and express their judgments (Bitektine, 2011; Bitektine & Haack, 2015; Tost, 2011). Known as the legitimacy-as-perception perspective (Suddaby et al., 2017), this approach conceptualizes legitimacy as residing "in the eye of the beholder" (Ashforth & Gibbs, 1990: 177) and focuses on the socio-cognitive processes that influence the legitimacy judgments of evaluators (Bitektine 2011; Gruban & Feix, 2024; Tost 2011). This shift highlights the importance of integrating individual-level dynamics to complement collective-level understandings of legitimacy. Our article adds to this micro-level perspective by examining how individual evaluators, such as consumers or citizens, form judgments and respond behaviorally to perceived (il)legitimacy.

At the individual level, evaluators' judgments are referred to as "propriety judgments," which reflect their judgment of whether a legitimacy object is appropriate for its social context (Tost, 2011). Scholars have developed context-specific scales to measure approval, acceptance, or agreement in individual-level surveys (e.g., Alexiou & Wiggins, 2019; Bitektine et al., 2020). Recent empirical research has used these measures to advance the understanding of the formation and expression of propriety judgments (Bitektine & Song, 2023; Zhang et al., 2022).

While there has been growing interest in the study of legitimacy judgments, comparatively little attention has been devoted to the behavioral manifestations of legitimacy. Building on the legitimacy-as-perception perspective and taking a step toward a behavioral perspective on legitimacy, we investigate the behavioral consequences of propriety judgments, which are assumed to guide evaluators' support or opposition to legitimacy objects (Bitektine, 2011; Hoefer & Green, 2016; Tost, 2011). To this end, and to circumvent issues such as hypothetical bias (Baumeister, Vohs, & Funder, 2007; Harrison & Rutström, 2008), we examine individuals' behaviors in an incentivized and controlled experimental setting (Hertwig & Ortmann, 2001). Our design allows us to observe whether evaluators' propriety judgments align with costly behavioral responses, thereby providing stronger empirical evidence of the behavioral manifestations of legitimacy.

HYPOTHESIS DEVELOPMENT

To examine the consistency between propriety judgments and behavioral responses, we use DeScioli and Kurzban's (2009) paradigm of moral interactions. This paradigm is particularly useful for experimental legitimacy research because of the important role that morals play in the conceptualization of legitimacy (Deephouse et al., 2017; Suchman, 1995; Suddaby et al., 2017). Indeed, a crucial determinant of legitimacy is the moral status of the entity, or the extent to which the legitimacy object conforms to evaluators' moral values (Deephouse & Suchman, 2008; Scott, 2001; Suchman, 1995). This focus on judgments about whether an action, practice, or organization is socially appropriate and conforms to normative expectations distinguishes legitimacy from other social evaluations, such as reputation and status (Bitektine, 2011; Deephouse & Carter, 2005).

The paradigm, illustrated in Figure 1, describes a strategic interaction in which a decision-maker (1) makes a decision that impacts both their own utility and the utility of a second party (2). This decision falls into the category of a moral decision because the decision-

maker can prevent the second party from being harmed (Bentham, 1972). In this paradigm, harm is operationalized as a decrease in the second party's utility. A third party, the evaluator (3), observes the decision and assesses the appropriateness of the decision. Therefore, the allocation decision is the focal legitimacy object in this paradigm, and the primary unit of interest is the evaluator who assesses the propriety of the decision-maker's allocation decisions. Consistent with previous research (Bitektine, 2011; Hoefer & Green, 2016; Tost, 2011), the underlying assumption is that propriety judgments and behavioral responses are consistent. That is, we expect to find effects in the same direction: if an evaluator judges a legitimacy object to be more proper, then it should receive more support. Conversely, if it is judged to be less proper, evaluators should be less willing to support it. However, we develop separate hypotheses for the evaluative and behavioral manifestations of legitimacy because they are measured by different outcomes. Figure 2 summarizes our theoretical model and hypotheses.

Insert Figure 1 and Figure 2 about here

Increasing Self-Interest

Evaluators' instrumental considerations (Tost, 2011) or "self-interested calculations" (Suchman, 1995: 578) are a key antecedent of propriety judgments. More specifically, evaluators assess whether a legitimacy object yields personal benefits such as monetary rewards (Suchman, 1995; Tost, 2011). In the context of the moral interaction paradigm (see Figure 1), increasing self-interest of the evaluator worsens the underlying trade-off (see Figure 1): both the evaluator and the decision-maker can increase their utility at the expense of the second party. While there are many situations in which evaluators can benefit without harming others, we focus on the more extreme case in which self-interest directly involves harming another party. Such situations exemplify corruption mechanisms driven by monetary incentives, wherein evaluators prioritize personal gain over the well-being of others (Frank &

Schulze, 2000; Lange, 2008). A pertinent example is the tobacco industry, which is widely perceived as more immoral than other industries due to its contribution to public health harms (Schneider, Brun, & Weber, 2020). Employees in this industry receive higher wages, which are understood to compensate for the moral costs associated with their work (Schneider et al., 2020). Building on this logic, we expect that if a legitimacy object generates personal gain for evaluators, even at the expense of others, they will be more likely to judge it as proper.

> *Hypothesis 1a:* Benefiting from a legitimacy object at the expense of a second party has a positive effect on evaluators' propriety judgments about the legitimacy object.

Once evaluators have formed a propriety judgment, this judgment guides behavior toward the legitimacy object (Tost, 2011: 697). Evaluators face three behavioral options: first, to the extent that the legitimacy object is viewed as proper, evaluators actively support it. Second, to the extent that the legitimacy object is viewed as improper, evaluators sanction it (Bitektine, 2011; Tost, 2011). Finally, propriety judgments about the legitimacy object may remain undetermined, meaning that evaluators neither actively support nor sanction the legitimacy object (Bitektine, 2011; Siraz et al., 2023).

Evaluators who benefit from a legitimacy object are motivated to support it (Tyler, 1997). In behavioral ethics, personal gain has been shown to lead individuals to disregard the unfavorable outcomes for a second party (Bazerman & Sezer, 2016; Chugh & Bazerman, 2007) and further push them to increase their wealth at the expense of others (Légeret & Zehnder, 2022). In fact, there is evidence that financial gain can lead people to give in to the temptation to deviate from their moral beliefs (Effron & Helgason, 2023). Therefore, we expect that evaluators who benefit from a legitimacy object that is unfavorable to the second party will provide more behavioral support for the legitimacy object than evaluators who do not benefit from an unfavorable legitimacy object.

Hypothesis 1b: Benefiting from a legitimacy object at the expense of a second party has a positive effect on evaluators' behavioral support for the legitimacy object.

Norm-Violating Legitimacy Object

Norms shape evaluators' beliefs about what constitutes an acceptable behavior and thus have a substantial impact on evaluators' propriety judgments (Aldrich & Fiol, 1994; Scott, 2001; Suchman, 1995). Norms generally prescribe or proscribe behaviors and can be defined as "shared understandings about actions that are obligatory, permitted, or forbidden" (Ostrom, 2000: 143–144; see also Elster, 1989). Propriety judgments thus depend on an assessment of whether the focal legitimacy object conforms to or violates the evaluator's conceptions of what is right and good, i.e., their normative expectations (Aldrich & Fiol, 1994; Suchman, 1995; Tost, 2011). Accordingly, a central consideration for evaluators is whether the focal legitimacy object is congruent with norms of appropriate behavior (Bitektine, 2011; Deephouse & Suchman, 2008). Congruence refers to the perceived alignment between the legitimacy object and norms (Schoon, 2022). Evaluators benchmark the focal legitimacy object against norms to determine whether the legitimacy object is acceptable, and therefore supported, or is unacceptable, and therefore sanctioned (Bitektine, 2011; Suchman, 1995; Tost, 2011).

Research in economics suggests that the perceived congruence of a decision with norms can be manipulated through the framing used to describe the decision (Krupka & Weber, 2013; List, 2007). Notably, Krupka and colleagues have shown that individuals evaluate decisions as more or less appropriate depending on the framing used to describe these decisions, even though the outcomes of these decisions are the same (Burks & Krupka, 2012; Krupka & Weber, 2013). Specifically, decisions that are framed as benevolent are likely to be perceived as norm-congruent and thus appropriate, whereas decisions that are framed as harmful are likely to be perceived as less appropriate because they violate the norm of doing no harm (Baron, 1996).

Therefore, we expect that an object that is framed as violating norms is judged as less appropriate (Bitektine & Haack, 2015).

Hypothesis 2a: Framing the legitimacy object as violating norms has a negative effect on evaluators' propriety judgments.

Research has shown that people are willing to punish actions that deviate from norms and reward actions that adhere to them (Fehr & Fischbacher, 2004; López-Pérez, 2008). For example, Fehr and Fischbacher (2004) found that in a dictator game, most people in the role of an evaluator choose to punish dictators who transfer less than half of their endowment to a second party. These results indicate that individuals engage in punishment when an observed behavior does not match their expectations of fairness and generosity. Therefore, we predict that evaluators will reduce their support for a legitimacy object when it is framed as violating norms compared to when it is framed as conforming to norms.

Hypothesis 2b: Framing the legitimacy object as violating norms has a negative effect on evaluators' behavioral support for the legitimacy object.

Norm Violation and Increased Self-Interest

While norm violation and self-interest can independently influence evaluators' propriety judgments and behaviors, their interaction creates a more complex evaluative dynamic, especially when normative and instrumental considerations are in conflict. For example, evaluators may view a legitimacy object that benefits them as proper on instrumental grounds, while at the same time viewing it as improper on normative grounds because they perceive it to violate norms (Suchman, 1995; Tost, 2011). This tension suggests that the combined influence of self-interest and norm violation on legitimacy evaluations is not trivial and may produce interaction effects that cannot be inferred from examining either factor in isolation (Gruban & Feix, 2024).

While the outcome of judgments that involve both considerations of personal gain and a norm violation is more difficult to predict, there is evidence that the monetary value tied to incentives can shift evaluators' propriety judgments, giving their self-interest priority over normative considerations (Aquino et al., 2009). Research on moral judgments has shown that individuals evaluate norm-violating behaviors as less improper when they serve their own interests, but they judge identical behaviors as improper when they do not benefit them personally. For example, Bocian and Wojciszke (2014) found that participants who observed another person cheat in an effort task to win a prize judged this behavior as less immoral when it allowed the observer to win an iPod than when the cheating did not result in a personal benefit for them. This finding suggests that high levels of personal gain can attenuate the negative effect of norm violation on judgments.

Similarly, Messick and Sentis (1979) suggest that people often compromise their moral standards when doing so leads to more personally beneficial outcomes, or may even ignore information about the negative consequences of their decisions altogether (Vu et al., 2023). This is consistent with findings that individuals tend to evaluate the actions of others as morally right when the expected outcomes are personally beneficial, and as morally wrong when they are personally harmful (Bocian, Baryla, & Wojciszke, 2020). Related findings indicate that people tend to evaluate a legitimacy object as more proper, despite a perceived norm violation, if it promotes their own self-interest (Foreman & Whetten, 2002; Tost, 2011). Therefore, we hypothesize that when a legitimacy object is framed as violating norms, evaluators who benefit from the legitimacy object will express more favorable propriety judgments than will evaluators who do not benefit from it. In other words, we expect self-interest to moderate the effects of norm violation on propriety judgments.

Hypothesis 3a: When a legitimacy object is framed as violating norms, the presence of self-interest (i.e., benefiting from the legitimacy object at the expense of a

second party) attenuates the negative effect of the norm-violating frame on evaluators' propriety judgments.

Individuals are more willing to engage in socially inappropriate behaviors when there is a personal benefit. For example, Aquino et al. (2009) found that participants in a negotiation task were more likely to lie when financial incentives rewarded negotiation outcomes. Similarly, individuals tend to misbehave when they can personally benefit from a norm-violating action. For example, people lie more when there are greater financial benefits to lying (Gneezy, 2005), and they spend less time working for charity when they can earn money for themselves instead (Légeret & Zehnder, 2022). These findings suggest that self-interest can override normative concerns, even in morally salient contexts. Therefore, we expect that when a legitimacy object is framed as violating norms, evaluators who benefit from it will show greater behavioral support than those who do not.

Hypothesis 3b: When a legitimacy object is framed as violating norms, the presence of self-interest (i.e., benefiting from the legitimacy object at the expense of a second party) attenuates the negative effect of the norm-violating frame on evaluators' behavioral support.

METHODS

Overview

We conducted two online experiments to examine how propriety judgments and behavioral support vary in response to norm congruence and self-interest. Both studies build on DeScioli and Kurzban's (2009) moral interaction paradigm. Each experimental group consisted of one Participant A (the decision-maker), who allocated a sum of money between themselves and a charity (serving as the second party), and ten Participants B (the evaluators), who observed the decision, choosing whether to reward or punish Participant A (see Figure 3).

Insert Figure 3 about here

Study 1 was conducted on MTurk and implemented all experimental manipulations. First, we varied whether Participant B's own interests were affected by Participant A's decision, thereby manipulating the presence or absence of self-interest. Second, we manipulated the framing of Participant A's decision by presenting it as either norm-congruent or norm-violating. Third, we varied the order of measurement by either asking Participants B to express a propriety judgment before engaging in a behavioral response or asking them to engage in a behavioral response without expressing a judgment. Study 2, conducted on Prolific, replicated the norm violation and propriety judgment manipulations using an independent U.S. sample. We recruited Participants A from individuals reporting managerial experience. Table 2 summarizes the key differences between the two studies.

Insert Table 2 about here

Methodological Rationale

The value of experimental research is increasingly recognized in organization and management studies (Bolinger et al., 2022; Levine et al., 2023), and particularly in legitimacy research (Haack et al., 2021; Suddaby et al., 2017). Our design responds to calls for more mechanism-based experimental approaches that provide a causally robust understanding of how legitimacy judgments and behaviors emerge (Bitektine & Haack, 2015; Deephouse et al., 2017). While experiments excel in internal validity by enabling the exogenous manipulation of key variables and rigorous testing of theoretical predictions (Greenberg & Tomlinson, 2004; Schram, 2005), they often raise concerns about generalizability to real-world settings (Highhouse, 2009; Scandura & Williams, 2000).

To balance internal validity with external realism, we conducted Study 1 on MTurk and Study 2 on Prolific, two platforms that provide access to a large and heterogenous pool of

participants and are widely used in management research (Aguinis et al., 2021; Hsu, Simmons, & Wieland, 2017; Shadish, Cook, & Campbell, 2002). We followed the best-practice recommendations for platform-based research to ensure data quality and ethical standards (Aguinis et al., 2021). Therefore, we designed the experiment to be concise and accessible. We included simple CAPTCHA tasks to promote engagement and screen out automated responses.

Design features ensuring behavioral authenticity. To maintain behavioral authenticity and avoid deception, we used the actual allocation decisions of Participants A rather than a fictional scenario. Using real decisions with monetary implications for a charitable organization minimizes hypothetical bias, that is, the tendency for stated intentions to diverge from actual actions in consequential situations (Baumeister et al., 2007; Harrison & Rutström, 2008). This bias can be particularly pronounced in moral contexts (Camerer & Mobbs, 2017).

We chose charitable giving as the allocation context because it reflects the common real-world trade-off between self-interest and prosocial concerns. The incentive structure ensured that increases in personal earnings for Participant A directly reduced donations to the charity, creating a clear and interpretable moral tension. This design choice enhances ecological validity while maintaining experimental control. Small donation decisions are common in daily life. For example, retailers often ask customers to round up their purchases to support a charitable cause (Neuman, 2024). These appeals typically ask for donations of less than $1.00 and provide minimal information about the beneficiary.

To capture the behavioral manifestations of legitimacy, we implemented an incentive-compatible reward and punishment mechanism. Participants B incurred real effort costs when rewarding or punishing Participant A, discouraging arbitrary responses and encouraging meaningful engagement. Although a single evaluator's action had modest impact, collective

punishment or reward could substantially affect Participant A's outcome. This reflects real-world collective dynamics, such as consumer boycotts or reputational sanctions.

Manipulating core antecedents of legitimacy. We manipulated self-interest by introducing financial incentives for Participants B, but only for allocation decisions favoring Participant A over the charity. This selective application of self-interest reflects real-world situations in which individuals are tempted to prioritize their own gain at the expense of others (Bentham, 1972). By structuring incentives in this way, we sought to capture the psychological and economic trade-offs involved in self-serving decisions. Moreover, we implemented this manipulation for half of the participants in each group to isolate the effect of self-interest while holding group-level resources constant, thereby reducing confounds related to wealth or endowment differences.

Norm congruence was manipulated through framing. Allocation decisions were described either as giving to a charity (norm-congruent) or taking from a charity (norm-violating), while holding material outcomes constant (Krupka & Weber, 2013; List, 2007). A parallel can be found in organ donation policies, where the framing of the decision significantly influences whether individuals choose to become donors. In opt-in systems, individuals must actively choose to donate their organs after death, framing the act as a voluntary donation. In contrast, opt-out systems assume that individuals are donors unless they choose to withdraw that consent, thus framing the act as taking back something that has already been designated for the public good. Although the ultimate outcomes (donating or not donating) are identical, donation rates are substantially higher in opt-out systems (Johnson & Goldstein, 2003).

Finally, to examine the relationship between the evaluative and behavioral manifestations of legitimacy, we implemented two measurement conditions. In one condition, Participants B expressed propriety judgments before engaging in behavioral responses. In the other condition, Participants B engaged only in behavioral responses. This design allowed us

to assess the alignment between judgments and behaviors while addressing concerns about order effects and endogeneity (Antonakis et al., 2010). Our approach is consistent with current research on the bidirectional relationship between judgments and behaviors, which emphasizes the importance of selecting a measurement order that aligns with the research question (Amasino, Pace, & van der Weele, 2024; Charness, Gneezy, & Rasocha, 2021).

Experimental Procedure

The experimental design was implemented in two steps. First, we collected the allocation decisions of Participants A. Second, we recruited participants from the same platform (Study 1: MTurk; Study 2: Prolific) to serve as Participants B. These participants were randomly assigned to the experimental conditions and exposed to the allocation decisions made by Participants A. Thus, the primary unit of analysis is the evaluators' (Participants B) propriety judgment and/or behavioral response to the allocation decisions.

Participants A. Participants A first completed ten CAPTCHA tasks (five characters), for which they received a fixed wage of $1.00. After completing the tasks, Participants A were asked to allocate a $1.00 bonus between themselves and a charity. They could choose from 11 discrete options in $0.10 increments (see Table 1). Options 1 to 6 favored the charity, whereas options 7 to 11 favored Participant A. Participants A were informed that their decisions would be evaluated by other participants, who could also reward or punish them, and that the resulting bonus would be paid after the study concluded.

Participants B. Participants B completed the same ten CAPTCHA tasks as Participants A for a fixed wage of $1.00. They were then randomly assigned to experimental conditions. Participants B were assigned into groups of ten and asked to vote for their preferred charity from a list of the five most popular charitable organizations in the U.S.: United Way Worldwide, Feeding America, Direct Relief, The Salvation Army, and St. Jude Children's Research Hospital. Each group was then randomly paired with one Participant A. To avoid

biasing Participants B's judgments and/or behaviors, the charity that received the most votes was not revealed until the end of the experiment.

Depending on their assigned condition, Participants B either (a) directly decided how they wished to reward or punish Participant A's decision, or (b) first evaluated the propriety of Participant A's decision before choosing a behavioral response. Using the strategy method (Brandts & Charness, 2011), Participants B provided their evaluative and/or behavioral responses for each of the 11 possible allocation decisions, without knowing which option Participant A had selected. The actual allocation was revealed at the end of the experiment, and the selected behavioral response was implemented.

Insert Table 1 about here

Measures and Experimental Manipulations

Behavioral support. We operationalized behavioral support as the decision by Participants B to reward or punish the allocation made by Participant A. Participants B could choose from five different degrees of support: A $0.10 punishment, a $0.05 punishment, neither a punishment nor a reward, a $0.05 reward, and a $0.10 reward. Participants B who chose the $0.10 punishment or reward had to complete two additional tasks. Participants B who chose to punish or reward $0.05 had to complete one additional task. Thus, rewarding or punishing an allocation decision came at a personal cost, requiring additional time and effort without increasing compensation, thereby encouraging deliberation and discouraging arbitrary responses.

Propriety judgments manipulation. To capture the effects of our treatments on propriety judgments and to test whether expressing judgments affected subsequent behaviors, we randomly allocated half of the participants to complete a propriety judgment questionnaire before making any behavioral decisions. This sequencing reflects the dominant theoretical

view in legitimacy research that judgments precede behavior (Bitektine, 2011; Tost, 2011). Consistent with prior work, we conceptualized propriety judgments as a continuous, bipolar construct, ranging from strongly positive to strongly negative, with a midpoint representing the threshold between propriety and impropriety (Siraz et al., 2023; Suddaby et al., 2017). We operationalized this construct on a seven-point Likert scale, ranging from "very inappropriate" to "very appropriate." We chose to operationalize propriety judgments as appropriateness because this concept captures the essence of legitimacy and is included in virtually all definitions of it (e.g., Schoon, 2022; Suchman, 1995; Tost, 2011). The centrality of appropriateness is also reflected in recent measurement approaches to legitimacy, which rely on evaluators' assessments of whether an object is appropriate, fair, or justified within a given social context (e.g., Schilke, Xue, & Haack, 2025). Thus, our operationalization is closely aligned with prevailing conceptualizations and measurement practices in the legitimacy literature.

Self-interest manipulation. We operationalized self-interest by attaching monetary incentives for Participants B to allocation decisions favoring Participant A over the charity (see Table 1, options 7 to 11). Within the group, half of Participants B received a fixed wage, independent of the allocation decision (control group), while the other half received a bonus that increased with the amount not allocated to the charity (self-interest).

Norm violation manipulation. Norm congruence was manipulated through framing. In the control group (norm-congruent framing), Participants A received a bonus and were asked to "give" money to a charity. In the norm-violating framing, a charity received the bonus and Participants A were asked to "take" money from a charity (Krupka & Weber, 2013; List, 2007). Financial outcomes were identical across framing conditions.

Detailed manipulation instructions are provided in Appendix A of the online supplement. The data and analysis codes for Study 1 and Study 2 are publicly available at https://osf.io/f2ps7/overview?view_only=3388c0a6e76c4b5eb4b485ecf269df6f.

STUDY 1

Data Collection and Demographics

We conducted our first study on MTurk using oTree (Chen, Schonger, & Wickens, 2016). First, we conducted a pretest using only the norm violation manipulation to measure its effect on propriety judgments and performed a power analysis to estimate the sample size required to replicate the findings of Krupka and Weber (2013). That is, we compared the propriety judgments observed in the control group and the norm violation treatment. The pretest indicated that, using a two-tailed alpha level of 0.05 and a power of 0.80, a sample size of approximately 190 participants per condition would be required to reach statistical significance. Accordingly, we continued data collection until we had at least 200 observations per condition to ensure sufficient statistical power to detect the hypothesized effects.

To ensure that participants were reliable and experienced, we restricted the sample to participants who had completed at least 500 tasks and had an approval rate of at least 95%. We also restricted participation to U.S. residents. First, we collected the decisions of 200 Participants A. Participants A took between three and seven minutes to complete the experiment, which translates to an hourly wage of between $8.75 and $19.50 per hour (not including possible rewards and punishments). Then, in a second step, we exposed 1,889 Participants B to Participant As' decisions. Participants B took between four and eight minutes to complete the experiment, which equates to an hourly wage of between $7.50 and $14.75 (not including bonuses). These wages are consistent with recommendations to pay participants fairly for their efforts, specifically above the U.S. federal minimum wage (Porter et al., 2019), which was $7.25 per hour at the time the study was conducted.

In our final sample, 54.69% of the participants identified as female. The median age of the participants falls between 36 and 40 years, and the median level of education is a bachelor's degree. While our primary analyses focus on estimating average treatment effects, detailed distributions of propriety judgments and behavioral support across all allocation options and treatment conditions are provided in Appendices B, C, and D of the online supplement.

Propriety Judgments

The regression results, presented in Table 3, include standard errors clustered at the individual level to account for the repeated-measures structure of the data. Because we used the strategy method, each Participant B provided responses for all 11 allocation options, resulting in 11 observations per participant. Figure 4 provides a visual summary of the behavioral support and highlights the main treatment effects.[1] We control for the allocation by using the 11 options as a continuous variable. We also control for whether or not the options favor the charity over Participant A, since the self-interest manipulation only affected allocations favoring Participant A. This allows us to further test whether our treatments have different effects depending on who is being favored by the decision.

Our results show that propriety decreased as less money was allocated to the charity in the control group (M = -0.142, SE = 0.009, $p < 0.001$), although there was no additional decrease for decisions favoring Participant A (M = 0.163, SE = 0.113, $p = 0.149$). This result is consistent with previous research on moral judgments showing that propriety judgments are sensitive to the outcome of the decision being evaluated (Malle, 2021). However, we find no support for Hypothesis 1a, as self-interest did not significantly lower propriety judgments for allocations favoring the charity (M = 0.032, SE = 0.133, $p = 0.811$), nor for allocations favoring Participant A (M = -0.048, SE = 0.142, $p = 0.733$). These results suggest that self-interest had no measurable effect on evaluators' propriety judgments.

Compared to the control group, the propriety judgments of decisions framed as norm-violating did not differ for options favoring the charity (M = -0.104, SE = 0.137, $p = 0.447$). However, the propriety judgments for options favoring Participant A were significantly different from one another (M = -0.449, SE = 0.153, $p = 0.004$). Decisions to take more than half of the money from the charity were evaluated as less appropriate than decisions not to give more than half of the bonus to the charity. Therefore, we find partial support for Hypothesis 2a, as decisions that were framed as norm-violating were judged as less proper for allocations that favored Participant A over the charity.

Finally, we find no support for Hypothesis 3a, as the combination of self-interest and norm-violating framing had no additional effect compared to norm-violating framing alone, for options favoring either the charity (M = -0.118, SE = 0.141, $p = 0.400$) or Participant A (M = -0.104, SE = 0.158, $p = 0.514$).

Insert Figure 4 and Table 3 about here

Behavioral Support Without Prior Measurement of Propriety Judgments

Table 4 summarizes the results of the regression with clustered standard errors at the individual level, and Figure 5 shows the average behavioral support for each allocation option, as well as a comparison of the treatments to their respective baselines. Consistent with the effects observed for propriety judgments, behavioral support decreased significantly as the amount being allocated to the charity decreased (M = -0.099, SE = 0.004, $p < 0.001$). However, there was an additional decrease in behavioral support for options that favored Participant A (M = -0.231, SE = 0.085, $p = 0.007$). We find no support for Hypothesis 1b, as self-interest had a significant negative impact on behavioral support for options favoring the charity (M = -0.244, SE = 0.090, $p = 0.007$). However, similar to the results for propriety judgments, self-

interest had no significant effect on behavioral support for options favoring Participant A (M = 0.032, SE = 0.110, *p = 0.774*).

Furthermore, we find support for Hypothesis 2b, as the decisions framed as violating norms received significantly less support than the control group for options favoring the charity (M = -0.272, SE = 0.088, *p = 0.002*) and for options favoring Participant A (M = -0.359, SE = 0.111, *p = 0.001*). Finally, we find no support for Hypothesis 3b, as the combination of self-interest and norm-violating framing had no significant effect on behavioral support compared to the norm-violating framing alone, for options favoring the charity (M = -0.108, SE = 0.092, *p = 0.237*) or options favoring Participant A (M = 0.214, SE = 0.118, *p = 0.070*).

Insert Figure 5 and Table 4 about here

Behavioral Support with Prior Measurement of Propriety Judgments

Although the treatment effects on propriety judgments and behavioral support appear to be aligned, it is possible that measuring propriety judgments affects subsequent behavioral responses (Siemsen, Roth, & Oliveira, 2010). To assess the impact of expressing propriety judgments on subsequent behavioral support, we compared the levels of behavioral support after expressing propriety judgments with those observed without it. Specifically, we examined the difference in treatment effects with and without prior measurement of propriety judgments. These results are reported in Table 4, along with the full regression for behavioral support. In addition, Figure 6 shows a comparison of the estimated treatment effects with and without prior measurement of propriety judgments. The average behavioral support for each allocation option across treatments is reported in Appendix E of the online supplement.

Our analysis revealed a notable difference in behavioral responses depending on whether participants had first expressed propriety judgments. Specifically, behavioral support for allocation decisions favoring the charity decreased significantly when participants had

previously evaluated the propriety of those decisions (M = -0.288, SE = 0.087, $p = 0.001$), whereas support for options favoring Participant A remained unaffected ($p = 0.687$). Participants B who expressed propriety judgments also showed significantly different reactions to self-interest for options favoring the charity (M = 0.353, SE = 0.124, $p = 0.005$) compared to those who reacted only behaviorally (through monetary rewards or punishments). For options favoring Participant A, however, self-interest effects were not significantly different (M = 0.096, SE = 0.146, $p = 0.511$).

In addition, framing decisions as norm-violating had significantly different effects on behavioral support for decisions favoring the charity (M = 0.380, SE = 0.123, $p = 0.002$) and for decisions favoring Participant A (M = 0.344, SE = 0.145, $p = 0.018$). However, the combined effects of self-interest and norm-violating framing were not significantly different for either set of options (M = -0.341, SE = 0.128, $p = 0.790$ and M = -0.165, SE = 0.158, $p = 0.296$, respectively).

Insert Figure 6 about here

STUDY 2

Data Collection and Demographics

To assess the robustness of our findings and to further examine the neutralization effect observed on MTurk, we conducted a preregistered replication study (OSF; DOI: 10.17605/OSF.IO/RKDTE) on Prolific with 801 U.S. participants. To maintain consistency with Study 1 conducted on MTurk, we restricted the sample to U.S. residents who had completed at least 500 tasks with an approval rate of at least 95%. We used the same financial stakes, instructions, and procedures as in Study 1. As a departure from Study 1, we conducted a 2×2 experiment, focusing on the Propriety judgments and Norm violation manipulations. This allowed us to test the robustness of the Norm violation manipulation and further

investigate the neutralization effect observed in Study 1. To enhance external validity, we departed from the MTurk design by recruiting Prolific participants who reported being currently employed in managerial positions as decision-makers (Participants A). Participants B were explicitly informed that these individuals held managerial roles in organizations. This adjustment reflects the organizational reality that managers often make allocation decisions with distributive implications. Therefore, it brings our experimental setup closer to contexts in which legitimacy judgments and audience responses unfold.

Similar to the MTurk data, the Prolific sample consisted of slightly more participants identifying as female (58.25% vs. 54.69% on MTurk). The median age of the Prolific participants was also higher (41-45 years vs. 36-40 years on MTurk), while the median level of education was comparable across both platforms (a bachelor's degree). To ensure comparability, we applied the same regression model as in the MTurk study, distinguishing between options favoring Participants A or the charity and clustering standard errors at the individual level.

Effects of Norm Violation on Propriety Judgments and Behavioral Support

Consistent with Study 1, both propriety judgments (M = -0.198, SE = 0.015, $p < 0.001$) and behavioral support (M = -0.127, SE = 0.007, $p < 0.001$) decreased as less money was allocated to the charity (see Figure 7 and Table 5). Furthermore, we replicated the partial support for Hypothesis 2a: propriety judgments for decisions framed as norm-violating differed for options favoring Participant A over the charity (M = -0.348, SE = 0.156, $p = 0.026$), but not for the options favoring the charity (M = -0.195, SE = 0.143, $p = 0.176$). These framing effects did not significantly differ from those observed in Study 1 for either options favoring the charity ($p = 0.648$) or options favoring Participant A ($p = 0.123$).

Furthermore, we found partial support for Hypothesis 2b, with a significant effect on behavioral support for options favoring Participant A (M = -0.275, SE = 0.127, $p = 0.030$), but

not for the other options (M = -0.101, SE = 0.098, $p = 0.302$). Unlike in Study 1, where the norm-violating framing affected all options, in Study 2 the effect was confined to those favoring Participant A. While the average treatment effects on behavioral support for options favoring the charity were not significantly different across the two platforms ($p = 0.190$), the effects for options favoring Participant A were significantly larger in Study 1 than in Study 2 ($p = 0.043$). The distributions of propriety judgments and behavioral support across all allocation options and treatment conditions are provided in Appendices F, G, and H of the online supplement.

Insert Figure 7 and Table 5 about here

Neutralization Effect

As illustrated in Figure 8 (see also Table 6), we found no evidence of a neutralization effect of the norm violating framing on behavioral support after the expression of propriety judgments for allocation options favoring the charity (M = -0.080, SE = 0.135, $p = 0.555$), nor Participants A (M = -0.047, SE = 0.170, $p = 0.782$). Compared to Study 1, the effects of expressing propriety judgments prior to behavioral responses differed significantly for options favoring the charity ($p = 0.012$), but not for options favoring Participant A ($p = 0.184$). Taken together, these findings suggest that, while the neutralization effect seems to depend on the evaluating audience, the central result, that norm violation systematically shapes legitimacy judgments and behavioral support, replicates across samples and platforms. We will discuss these findings further in the following section.

Insert Figure 8 and Table 6 about here

DISCUSSION AND CONCLUSION

Building on prior work that conceptualizes legitimacy as a perception, our article takes first steps toward advancing a behavioral perspective on legitimacy by examining when and why do legitimacy judgments correspond to behavior, and when do they not. To answer this question, we examined the consistency of evaluators' propriety judgments and behavioral responses, and how they are influenced by two key antecedents of legitimacy: self-interest and norm violation. Using controlled experimental designs, we compared behavioral support with and without the prior expression of propriety judgments across two samples.

Our findings reveal a nuanced pattern. Propriety judgments were largely unaffected by self-interest yet sensitive to norm violation, especially when the allocation decisions strongly favored Participant A over the charity. In contrast, behavioral responses varied across the full range of allocations, with lower support for norm-violating actions regardless of whether Participant A or the charity was favored. Self-interest reduced behavioral support for allocations favoring the charity, suggesting that personal incentives influenced behavior even when propriety judgments remained unaffected. Among MTurk participants, expressing judgments attenuated subsequent behavioral responses, an effect not observed on Prolific.

Together, these results offer a comprehensive understanding of the conditions under which legitimacy judgments correspond with behavioral support. The next sections discuss the implications of these results for research on legitimacy and social evaluations.

The (In)Consistency of Propriety Judgments and Behaviors

Research grounded in the legitimacy-as-perception perspective assumes that legitimacy manifests itself in consistent judgments and behaviors: "on the one hand, to the extent that an entity is viewed as legitimate, it is supported, and attempts to change it are resisted; on the other hand, to the extent that an entity is viewed as illegitimate, people actively seek to change it" (Tost, 2011: 697). In real life, however, judgments and behaviors often diverge. Consumers

often express moral disapproval of fast-fashion brands yet continue to support them through their purchasing behavior (McKinsey, 2025). Similarly, many people express concern about climate change, yet they persist in high-carbon practices that damage the environment, such as frequent air travel (Árnadóttir, Czepkiewicz, & Heinonen, 2021). These examples suggest that expressed evaluations do not automatically translate into action. Despite the centrality of this assumption in legitimacy research, there is little empirical evidence, particularly from controlled experiments, examining whether and when propriety judgments correspond to actions.

Our results show that the relationship between legitimacy judgments and behavioral support is more complex than previously theorized. Decisions that conformed to norms (i.e., "giving" money to a charity) were evaluated as more appropriate and received significantly more support than decisions that violated norms (i.e., "taking" from a charity). That is, the effect of norm violation was consistent across judgments and behaviors. This finding is robust across samples and is consistent with the legitimacy-as-perception perspective (Bitektine, 2011; Bitektine & Haack, 2015; Tost, 2011). However, propriety judgments and behavioral responses can be inconsistent, particularly when personal gain is involved: whereas judgments remained insensitive, behaviors were influenced by financial incentives.

Although much of the literature treats judgments and behaviors as closely aligned, prior work has noted that legitimacy processes can be fragmented and inconsistent across different audiences (Bitektine & Haack, 2015; Suddaby et al., 2017). Our findings empirically substantiate this view, showing that legitimacy manifests in multiple, and sometimes contradictory, ways through individual judgments and behaviors. These findings reinforce the broader theoretical understanding of legitimacy as a complex, dynamic, and context-sensitive phenomenon. This pattern is also evident in real-world cases where expressed judgments and actions both converge and diverge.

A possible explanation for the inconsistencies between propriety judgments and behaviors observed for the exposure to personal gain is the gap between intentions and actual behavior (Ajzen et al., 2004; Sheeran & Webb, 2016). The experimental results from Study 1 highlight the specific conditions under which this gap emerges in the context of legitimacy: evaluators may express favorable or unfavorable propriety judgments yet act in ways that deviate from these evaluations when given the opportunity to do so. These findings caution against assuming that behavioral support can be reliably inferred from self-reported judgments alone. This concern is particularly relevant for survey-based research, where participants' stated evaluations or intentions are often treated as proxies for actual behavior (Baumeister et al., 2007). Such measures are susceptible to social desirability bias, which can lead to distorted or inconsistent responses (Banks, Woznyj, & Mansfield, 2023). Consequently, existing measurement instruments of legitimacy (e.g., Alexiou & Wiggins, 2019; Bitektine et al., 2020) should be supplemented with behavioral measures to provide a more comprehensive assessment.

Furthermore, judgments are often not costly in the context of an experiment or survey, allowing evaluators to signal insensitivity to self-interest and maintain a positive self-image in moral contexts (Banaji, Bazerman, & Chugh, 2003). However, research indicates that individuals frequently overestimate their own morality (Epley & Dunning, 2000), as evidenced by Sedikides et al.'s (2014) finding that prison inmates believe they are more moral than the average person outside of prison. Our findings suggest that, although propriety judgments appear unaffected by self-interest, actual behaviors tell a different story. When behaviors have real consequences, evaluators often prioritize personal gain, even if it harms a charitable cause, revealing the latent influence of self-interest that judgments alone may obscure.

Together, Studies 1 and 2 distinguish what is robust from what is contingent. Across both platforms, norm-violating framing systematically shifted propriety judgments and reduced

behavioral support, indicating that norm congruence is a reliable driver of evaluative and behavioral manifestations of legitimacy. At the same time, the two studies diverge in how strongly audience members translated their judgments into costly actions and, in particular, whether expressing a judgment altered subsequent behavior. This cross-study pattern suggests that while the directional role of norm violation is stable, the link between judgment and action is context-sensitive. We address this possibility in the next section.

Overall, our findings suggest that evaluators' propriety judgments and behaviors are less consistent than the legitimacy-as-perception perspective often assumes. This insight serves as an important caution for scholars: inferring consistent behavioral support from expressed propriety judgments may lead to inaccurate theoretical conclusions and misguided policy recommendations. Our findings underscore the need for legitimacy research to distinguish between evaluative and behavioral manifestations. Theoretical frameworks and empirical studies alike must recognize this distinction to ensure that conclusions are both robust and actionable (Banks et al., 2023).

Does Voicing a Propriety Judgment Neutralize Action?

In Study 1 on MTurk, the significant effects of norm violation as well as self-interest on participants' behavioral support for the legitimacy object diminished or completely disappeared when evaluators were asked to express their judgments before engaging in behavioral responses. In Study 1, measuring propriety judgments prior to behavioral support attenuated, or eliminated in some cases, the behavioral effects of both norm violation and self-interest. This indicates that the role of propriety judgments is ambivalent. On the one hand, articulating a negative judgment may mitigate the influence of self-interest, potentially acting as a commitment device that strengthens moral consistency. On the other hand, such expressions may also reduce the likelihood of punishing norm-violating actions, possibly

because evaluators feel they have already done their part by judging the behavior to be inappropriate.

One possible explanation for why the *neutralization* occurs is that evaluators have already expressed their propriety judgments. When individuals evaluate a legitimacy object as inappropriate, articulating a negative judgment may relieve negative affect and reduce the perceived need for further action. For example, Xiao and Houser (2005) showed that expressing negative emotions can reduce the likelihood of engaging in costly punishment.

An alternative explanation is that expressing a judgment toward a legitimacy object that is perceived as inappropriate may help evaluators maintain a positive self-image, even if they do not act in accordance with that judgment. By voicing a negative judgment, evaluators symbolically adhere to their moral principles without incurring the cost of sanctioning the legitimacy object. This process reflects a form of strategic response commonly referred to as motivated reasoning (Zimmermann, 2020). Through motivated reasoning, people can reconcile inconsistencies between their moral judgments and behaviors by selectively construing their inaction as still morally justifiable (Haisley & Weber, 2010).

A related mechanism is moral licensing, which occurs when individuals perceive that expressing moral intentions or values entitles them to act less virtuously afterward (Effron & Monin, 2010). In this view, expressing disapproval of a norm violation can serve as a form of moral "credit," thereby reducing internal pressure to engage in costly or effortful action. The act of expressing a judgment can also reduce cognitive dissonance by aligning moral expression with one's self-concept, which in turn reduces the felt need to take further action. Together, these psychological mechanisms may help explain why the prior expression of propriety judgments attenuates the behavioral effects of norm violation and self-interest in our study.

However, in Study 2 on Prolific, we did not observe the neutralization effect: behavioral support remained sensitive to norm violation regardless of whether judgments were elicited

beforehand. The contrast between Study 1 and Study 2 provides useful diagnostic insight into neutralization. Two differences are particularly relevant. First, Study 2 introduced a managerial decision-maker. This may cause evaluators to assign more responsibility to the decision-maker and increase the meaningfulness of sanctioning, making "voicing disapproval" seem less psychologically sufficient. Second, Prolific's stronger platform norms and higher data quality expectations may foster greater task engagement, reducing the tendency to substitute "cheap" moral expression for costly action. While our design does not isolate these explanations definitively, the combined evidence suggests that neutralization is not a universal consequence of measuring propriety judgments. Rather, it is a context-sensitive process that emerges when moral expression can plausibly substitute for action. This insight aligns with the emphasis on heterogeneity in audiences and evaluation criteria in legitimacy research (Suddaby et al., 2017; Lamin & Zaheer, 2012; Siraz et al., 2023). Overall, our findings suggest that propriety judgments can function as both *cheap talk,* costless moral signaling, and as genuine moral commitment. Whether they serve one or the other role may depend on the social context and the perceived sufficiency of the expression as a moral act.

We believe that our results point to a complex yet realistic pattern in how people act on their legitimacy judgments. On the one hand, there are cases of consistent evaluative and behavioral manifestations, such as the Volkswagen emissions scandal. On the other hand, there are cases of inconsistent evaluative and behavioral manifestations. Apart of the example of the fast fashion industry, another real-world example of how the neutralization effect might play out can be seen in the case of the 2022 FIFA World Cup in Qatar. Prior to the tournament, several European soccer federations, officials, and players publicly criticized Qatar's human rights record, particularly regarding the treatment of migrant workers and the country's stance on LGBTQ+ rights (Mandard, Dupré, & Barthe, 2022). The Dutch soccer federation, for example, stated that it has "never been in favour of holding the World Cup in Qatar and of

course certainly doesn't approve of the way in which migrant workers are treated there" (KNVB, n.d.). Despite such evaluative statements, the European federations decided to participate in the tournament, refusing to act on calls for a boycott (Mac Dougall, 2022). This divergence between expressed negative legitimacy judgments and inaction may reflect the dynamic identified in Study 1: the expression of judgment may reduce the perceived need for further behavioral action.

Managerial Implications

In addition to contributing to legitimacy research and social evaluation research more broadly, our findings provide valuable insights for managerial practice. Managers should be aware of the potential misalignment between the judgments of organizational members and their behaviors, particularly in contexts involving financial incentives. This discrepancy suggests that individuals may act inconsistently with their expressed judgments, especially when tempted by personal gain. Expressed judgments may serve as a commitment device, helping to guard against unethical behavior. Organizations might therefore consider interventions that encourage employees to express their concerns, as such practices could mitigate the effects of temptation on behavior. For example, structured forums or anonymous feedback mechanisms could allow employees to express ethical concerns before acting. Nevertheless, practitioners should also be aware that voicing mechanisms could inadvertently suppress desirable actions, such as taking initiative to address organizational challenges. Therefore, these tools should be used judiciously, focusing on situations where the behavioral consequences of temptation are clearly harmful. Overall, our findings underscore the importance of raising awareness of both the judgment-behavior gap and the influence that expressed judgments can have on behavior. This insight can guide the development of training programs, decision-making protocols, and ethical safeguards to foster more consistent and constructive organizational behavior.

Methodological Implications for Legitimacy Research

While the possibility of neutralizing action through judgment expression is a theoretically intriguing finding, it presents significant methodological challenges to scholars who seek to measure both legitimacy judgments and behaviors within the same study. Our results suggest that the very act of eliciting propriety judgments may alter subsequent behavioral responses, thereby biasing the outcomes and potentially masking important treatment effects. A key implication for future experimental research on legitimacy is the need to study judgments and behaviors separately to avoid biasing behaviors with the prior expression of judgments. Including both judgments and behavioral measures not only introduces the potential for neutralization effects observed in our study, which may obscure the true effects of treatments on behaviors, but also raises potential endogeneity concerns (Antonakis et al., 2010). Specifically, measuring multiple variables from the same source, i.e., the same individual, can lead to endogeneity issues due to shared variance between these variables, as has been well-documented in prior research (Siemsen et al., 2010). As a result, researchers interested in studying behaviors should focus on measuring actual behaviors, whereas those studying judgments should rely on measures that capture evaluations independently (Amasino et al., 2024; Charness et al., 2021). For scholars aiming to examine both judgments and behaviors, it is essential to measure them separately to prevent one from influencing the other (d'Adda et al., 2016; Rustichini & Villeval, 2014).

Moreover, we encourage researchers to incentivize their measures and opt for real, effortful punishments and rewards when studying how people behave toward legitimacy objects. Unincentivized measures, such as hypothetical decisions, are typically associated with imaginative processes, whereas incentivized behaviors activate both affective and social cognitive processes (FeldmanHall et al., 2012). As a result, relying on hypothetical behaviors or intentions can introduce a hypothetical bias—the gap between hypothetical and actual

decisions. This bias is particularly concerning in the context of moral decision-making, as individuals often overestimate their own moral standards (Camerer & Mobbs, 2017; Légeret & Gruban, 2024). Our results demonstrate that it is possible to capture the effects of key antecedents of legitimacy on costly behaviors. Despite the personal cost associated with punishing or rewarding the legitimacy object, participants engaged in both types of actions, responding systematically to our experimental treatments. These findings underscore the feasibility and value of studying real, incentivized behaviors in legitimacy research. By moving beyond hypothetical scenarios and self-reported intentions, our study offers a promising foundation for future investigations of the behavioral manifestations of legitimacy, advancing both methodological rigor and theoretical insights in this important area of organization and management research.

Limitations and Avenues for Future Research

Despite our efforts to ensure mundane realism (Berkowitz & Donnerstein, 1982; Greenberg & Tomlinson, 2004), that is, to design the experiment to resemble real-life situations, our research still has limitations with respect to external validity. As with most experimental studies, questions remain about the generalizability of the findings beyond the controlled setting. While our approach prioritizes internal validity and causal identification, we believe that future studies could meaningfully extend this research by adapting our paradigm to field settings. Doing so would enhance ecological validity and further our understanding of how legitimacy judgments and behaviors unfold in real organizational and institutional environments.

At the same time, Study 2 suggests that platform and sample characteristics may influence findings. Specifically, we consistently observed a neutralization effect on MTurk, but not on Prolific. While this suggests that neutralization may be context-dependent rather than universal, it raises important questions about how differences in participant pools and the

social context shape legitimacy evaluations and behavioral responses (Levitt & List, 2007). We therefore encourage future work to systematically compare different participant sources and to extend replications to the field where consequential organizational decisions are made by employees and managers.

Another limitation of our research is that participants interacted anonymously and without preexisting social relationships. While most experiments ensure participant anonymity, the use of online platforms such as MTurk or other crowdsourcing platforms, where participants are recruited from a heterogeneous pool, can amplify the lack of shared identity. Participants interacted just once, and the only information they had about each other was their mutual participation in the same study. However, legitimacy inherently involves a relationship between an evaluator and a legitimacy object (Suchman, 1995), along with certain expectations that define the nature of that relationship (Schoon, 2022). With that in mind, a fruitful avenue for future research is to introduce social interactions between participants to examine the inherently relational (Tost, 2011) and communicative (Haack et al., 2021) dimensions of legitimacy. Such designs would not only enhance the external validity of experimental studies but also allow for the investigation of the role of validity cues, such as peer endorsement or authorization by high-status actors. Such cues merit further attention because legitimacy does not occur in a social vacuum, but involves persuasion and social influence (Johnson, Dowd, & Ridgeway, 2006). Exploring the behavioral manifestations of legitimacy in more socially interactive settings, whether in controlled laboratory experiments, field studies, or natural experiments, would provide valuable insights into the dynamics of legitimacy and its broader implications.

Another promising research avenue involves the cross-cultural study of the evaluative and behavioral manifestations of legitimacy. That is, it would be particularly valuable to examine whether the neutralization effect or punishments for norm-violating actions are

consistent across cultural contexts. For example, previous research has shown that unequal outcomes are met with various degrees of punishment across cultures (Henrich et al., 2006). We propose that other-orientation—the extent to which individuals consider the well-being of others (de Dreu & Nauta, 2009)—that is known to vary across cultures (Meglino & Korsgaard, 2004) may have a substantial impact on both legitimacy judgments and behaviors.

We have relied on costly behaviors for our main dependent variable. Despite evidence suggesting that MTurk participants are primarily motivated by maximizing their earnings (Aguinis et al., 2021; Hauser & Schwarz, 2016), the participants who served as evaluators in our experiment voluntarily spent additional time, thereby reducing their hourly wage, to punish or reward the allocation decisions made by Participants A. Study 2 revealed similar patterns. Based on the idea that evaluators reward legitimacy objects they deem appropriate and sanction those they consider inappropriate (Bitektine, 2011; Tost, 2011), making support for allocation decisions costly increases the external validity of our experiment. Accordingly, future legitimacy research should prioritize costly behavioral measures because only decisions that entail real trade-offs can credibly capture the behavioral manifestations of legitimacy.

Finally, future research can fruitfully explore the circumstances under which the neutralization effect becomes more pronounced or diminishes. For example, we expect neutralization to be more likely in contexts where behavioral support is costly—such as when it requires significant effort, time, or resources—while the expression of judgments remains relatively cost-free. In contrast, we expect neutralization to be less likely to occur when judgments and behaviors are made publicly or are observed by others, as individuals may strive to avoid being perceived as inconsistent or hypocritical. The results of our two studies illustrate the complexity of this phenomenon: while the effects of norm violation on propriety judgments and behaviors appear consistent across audiences, the extent to which judgment expression attenuates behavior varies across context. Examining these dynamics in both controlled

laboratory settings and field studies, particularly in high-stakes environments where social norms are salient, can shed light on the factors that drive the alignment—or misalignment—between propriety judgments and behaviors. Such research would advance our understanding of how legitimacy is enacted, signaled, and reinforced in organizational and institutional contexts, and deepen theoretical perspectives on the role of social context in shaping moral behavior. We hope that future research will build on these insights to further develop and extend a behavioral perspective of legitimacy, clarifying when legitimacy is merely expressed and when it is actively enacted through costly behavior.

REFERENCES

Aguinis, H., Villamor, I., & Ramani, R. S. 2021. MTurk research: Review and recommendations. *Journal of Management*, 47: 823-837.

Ajzen, I., Brown, T. C., & Carvajal, F. 2004. Explaining the discrepancy between intentions and actions: The case of hypothetical bias in contingent valuation. *Personality and Social Psychology Bulletin*, 30: 1108-1121.

Aldrich, H. E., & Fiol, C. M. 1994. Fools rush in? The institutional context of industry creation. *Academy of Management Review*, 19: 645-670.

Alexiou, K., & Wiggins, J. 2019. Measuring individual legitimacy perceptions: Scale development and validation. *Strategic Organization*, 17: 470-496.

Amasino, D. R., Pace, D. D., & van der Weele, J. J. 2024. Fair shares and selective attention. *American Economic Journal: Microeconomics*, 16: 259-290.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. 2010. On making causal claims: A review and recommendations. *Leadership Quarterly*, 21: 1086-1120.

Aquino K., Freeman, D., Reed, A., Felps, W., & Lim, V. K. 2009. Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, 97: 123-141.

Árnadóttir, Á., Czepkiewicz, M., & Heinonen, J. 2021. Climate change concern and the desire to travel: How do I justify my flights? *Travel Behaviour and Society,* 24: 282-290.

Ashforth, B. E., & Gibbs, B. W. 1990. The double-edge of organizational legitimation. *Organization Science*, 1: 177-194.

Banaji, M. R., Bazerman, M., & Chugh, D. 2003. How (un)ethical are you? *Harvard Business Review*, 81: 56-64.

Banks, G. C., Woznyj, H. M., & Mansfield, C. A. 2023. Where is "behavior" in organizational behavior? A call for a revolution in leadership research and beyond. *The Leadership Quarterly*, 34: 1-15.

Baron, J. 1996. Do no harm. In D. M. Messick & A. E. Tenbrunsel (Eds.), *Codes of conduct: Behavioral research into business ethics*: 197-213. New York: Russell Sage Foundation.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. 2007. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2: 396-403.

Bazerman, M. H., & Sezer, O. 2016. Bounded awareness: Implications for ethical decision making. *Organizational Behavior and Human Decision Processes*, 136: 95-105.

Bentham, J. 1972. *An introduction to the principles of morals*. London: Athlone.

Bergh, D. D., Boyd, B. K., Byron, K., Gove, S., Ketchen, D. J. 2022. What constitutes a methodological contribution? *Journal of Management*, 48: 1835-1848.

Berkowitz, L., & Donnerstein, E. 1982. External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist*, 37: 245-257.

Bitektine, A. 2011. Toward a theory of social judgments of organizations: The case of legitimacy, reputation, and status. *Academy of Management Review*, 36: 151-179.

Bitektine, A., & Haack, P. 2015. The "macro" and the "micro" of legitimacy: Toward a multilevel theory of the legitimacy process. *Academy of Management Review*, 40: 49-75.

Bitektine, A., Hill, K., Song, F., & Vandenberghe, C. 2020. Organizational legitimacy, reputation, and status: Insights from micro-level measurement. *Academy of Management Discoveries*, 6: 107-136.

Bitektine, A., & Song, F. 2023. On the role of institutional logics in legitimacy evaluations: The effects of pricing and CSR signals on organizational legitimacy. *Journal of Management*, 49: 1070-1105.

Bocian, K., & Wojciszke, B. 2014. Self-interest bias in the moral judgments of others' actions. *Personality and Social Psychology Bulletin*, 40: 898-909.

Bocian, K., Baryla, W., & Wojciszke, B. 2020. Egocentrism shapes moral judgements. *Social and Personality Psychology Compass*, 14: 1-14.

Bolinger, M. T., Josefy, M. A., Stevenson, R., & Hitt, M. A. 2022. Experiments in strategy research: A critical review and future research opportunities. *Journal of Management*, 48: 77-113.

Brandts, J., & Charness, G. 2011. The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, 14: 375-398.

Burks, S. V., & Krupka, E. L. 2012. A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science*, 58: 203-217.

Camerer, C., & Mobbs, D. 2017. Differences in behavior and brain activity during hypothetical and real choices. *Trends in Cognitive Sciences*, 21: 46-56.

Charness, G., Gneezy, U., & Rasocha, V. 2021. Experimental methods: Eliciting beliefs. *Journal of Economic Behavior & Organization*, 189: 234-256.

Chen, D. L., Schonger, M., & Wickens, C. 2016. oTree – An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9: 88-97.

Chugh, D., & Bazerman, M. H. 2007. Bounded awareness: what you fail to see can hurt you. *Mind & Society*, 6: 1-18.

d'Adda, G., Drouvelis, M., & Nosenzo, D. 2016. Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, 62: 1-7.

de Dreu, C. K. W., & Nauta, A. 2009. Self-interest and other-orientation in organizational behavior: Implications for job performance, prosocial behavior, and personal initiative. *Journal of Applied Psychology*, 94: 913-926.

Deephouse, D. L., & Carter, S. M. 2005. An examination of differences between organizational legitimacy and organizational reputation. *Journal of Management Studies*, 42: 329-360.

Deephouse, D. L., Bundy, J., Tost, L. P., & Suchman, M. C. 2017. Organizational legitimacy: Six key questions. In R. Greenwood, C. Oliver, T. B. Lawrence & R. E. Meyer (Eds.), *The Sage handbook of organizational institutionalism* (2nd ed.): 27-54. Thousand Oaks, CA: Sage.

Deephouse, D. L., & Suchman, M. C. 2008. Legitimacy in organizational institutionalism. In R. Greenwood, C. Oliver, K. Sahlin, & R. Suddaby (Eds.), *The Sage handbook of organizational institutionalism*: 49-77. Thousand Oaks, CA: Sage.

DeScioli, P., & Kurzban, R. 2009. Mysteries of morality. *Cognition*, 112: 281-299.

Effron, D. A., & Helgason, B. A. 2023. Moral inconsistency. *Advances in Experimental Social Psychology*, 67: 1-55.

Effron, D. A., & Monin, B. 2010. Letting people off the hook: When do good deeds excuse transgressions? *Personality and Social Psychology Bulletin*, 36: 1618-1634.

Elster, J. 1989. Social norms and economic theory. *The Journal of Economic Perspectives*, 3: 99-117.

Epley, N., & Dunning, D. 2000. Feeling "holier than thou": Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology*, 79: 861-875.

Fehr, E., & Fischbacher, U. 2004. Third-party punishment and social norms. *Evolution and Human Behavior*, 25: 63-87.

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. 2012. Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social Cognitive and Affective Neuroscience*, 7: 743-751.

Foreman, P., & Whetten, D. A. 2002. Member's identification with multiple-identity organizations. *Organization Science*, 13: 618-635.

Frank, B., & Schulze, G. G. 2000. Does economics make citizens corrupt? *Journal of Economic Behavior & Organization*, 43: 101-113.

Gneezy, U. 2005. Deception: The role of consequences. *American Economic Review*, 95: 384-394.

Greenberg, J., & Tomlinson, E. C. 2004. Situated experiments in organizations: Transplanting the lab to the field. *Journal of Management*, 30: 703-724.

Greenwood, R., Oliver, C., Lawrence, T. B., & Meyer, R. E. 2017. Introduction: Into the fourth decade. In R. Greenwood, C. Oliver, T. B. Lawrence & R. E. Meyer (Eds.), *The Sage handbook of organizational institutionalism* (2nd ed.): 1-24. Thousand Oaks, CA: Sage.

Gruban, M., & Feix, A. in press. Facilitating marketization *à contrecœur*: Why stakeholders may continue to support organizations that introduce market practices violating their values. *Journal of Management Studies.* doi:10.1111/joms.13158.

Haack, P., Schilke, O., & Zucker, L. G. 2021. Legitimacy revisited: Disentangling propriety, validity, and consensus. *Journal of Management Studies*, 58: 749-781.

Haisley, E. C., & Weber, R. A. 2010. Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior*, 68: 614-625.

Harrison, G. W., & Rutström, E. E. 2008. Experimental evidence on the existence of hypothetical bias in value elicitation methods. *Handbook of Experimental Economics Results*, 1: 752-767.

Hauser, D. J., & Schwarz, N. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48: 400-407.

Henrich, J., McElreath, R., Barr, A., Ensimger, J., Barrett, C., Bolyanatz, A., et al. 2006. Costly punishment across human societies. *Science*, 312: 1767-1770.

Hertwig, R., & Ortmann, A. 2001. Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24: 383-451.

Highhouse, S. 2007. Designing experiments that generalize. *Organizational Research Methods*, 12: 554-566.

Hoefer, R. L., & Green, S. E., Jr. 2016. A rhetorical model of institutional decision making: The role of rhetoric in the formation and change of legitimacy judgments. *Academy of Management Review*, 41: 130-150.

Hsu, D. K., Simmons, S. A., & Wieland, A. M. 2017. Designing entrepreneurship experiments: A review, typology, and research agenda. *Organizational Research Methods*, 20: 379-412.

Johnson, C., Dowd, T. J., & Ridgeway, C. L. 2006. Legitimacy as a social process. *Annual Review of Sociology*, 32: 53-78.

Johnson, E. J., & Goldstein, D. 2003. Do defaults save lives? *Science*, 302: 1338-1339.

Jolly, J. 2019. Volkswagen emissions scandal: mass lawsuit opens in Germany. *The Guardian*, Retrieved from https://www.theguardian.com/business/2019/sep/30/volkswagen-emissions-scandal-mass-lawsuit-opens-in-germany.

KNVB [Royal Dutch Football Association] (n.d.). The KNVB and the World Cup in Qatar. Retrieved from https://www.knvb.com/info/1151/knvb-and-world-cup-qatar.

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: 495-524.

Lamin, A., & Zaheer, S. 2012. Wall Street vs. Main Street: Firm strategies for defending legitimacy and their impact on different stakeholders. *Organization Science*, 23: 47-66.

Lange, D. 2008. A multidimensional conceptualization of organizational corruption control. *Academy of Management Review*, 33: 710-729.

Légeret, M., & Gruban, M. 2024. Are we as moral as we intend? Financial temptation, social norms, and the hypothetical bias. *Academy of Management Proceedings*, 10595.

Légeret, M., & Zehnder, C. 2022. Self-regulation after temptation? *Economics Letters*, 214: 110436.

Levine, S. S., Schilke, O., Kacperczyk, O., & Zucker, L. G. 2023. Primer for experimental methods in organization theory. *Organization Science*, 34: 1997-2525.

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

List, J. A. 2007. On the interpretation of giving in dictator games. *Journal of Political Economy*, 115: 482-493.

López-Pérez, R. 2008. Aversion to norm-breaking: A model. *Games and Economic Behavior*, 64: 237-267.

Mac Dougall, D. 2022. FIFA and Qatar 'rattled' as European football World Cup boycott gathers pace. *Euronews*. Retrieved from https://www.euronews.com/2022/11/10/fifa-and-qatar-rattled-as-european-world-cup-boycott-gathers-pace.

Malle, B. F. 2021. Moral judgments. *Annual Review of Psychology*, 72: 293-318.

Mandard, S., Dupré, R., & Barthe, B. 2022. Qatar World Cup: The reasons behind the widespread criticism. *Le Monde*, October 17. Retrieved from https://www.lemonde.fr/en/sports/article/2022/10/17/the-reasons-behind-the-widespread-criticism-of-qatar-s-world-cup_6000584_9.html.

McKinsey 2025. What is fast fashion? Retrieved from https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-fast-fashion.

Meglino, B. M., & Korsgaard, A. 2004. Considering rational self-interest as a disposition: Organizational implications of other orientation. *Journal of Applied Psychology*, 89: 946-959.

Messick, D. M., & Sentis, K. P. 1979. Fairness and preference. *Journal of Experimental Social Psychology*, 15: 418-434.

Neuman, S. 2024. That spare change you donate at checkout is adding up to millions for charities. *National Public Radio*, March 10. Retrieved from https://www.npr.org/2024/03/10/1236458377/charity-roundup-donations-stores-fundraising.

Ostrom, E. 2000. Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14: 137-158.

Porter, C. O. L. H., Outlaw, R., Gale, J. P., & Cho, T. S. 2019. The use of online panel data in management research: A review and recommendations. *Journal of Management*, 45: 319-344.

Rustichini, A., & Villeval, M. C. 2014. Moral hypocrisy, power and social preferences. *Journal of Economic Behavior & Organization*, 107(Pt A): 10-24.

Scandura, T. A., & Williams, E. A. 2000. Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, 43: 1248-1264.

Schilke, O., Xue, Z., Haack, P. 2025. Legitimacy construction in the presence of multiple validity cues: An experimental investigation. In J. E. Stets, K. A. Hegtvedt, & L. Doan (Eds.), *Handbook of social psychology. Handbooks of sociology and social research*. Springer, Cham.

Schneider, F., Brun, F., Weber, R. A. 2020. *Sorting and wage premiums in immoral work.* Working Paper no. 353, University of Zurich, Switzerland.

Schoon, E. W. 2022. Operationalizing legitimacy. *American Sociological Review*, 87: 478-503.

Schram, A. 2005. Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12: 225-237.

Scott, W. R. 2001. *Institutions and organizations* (2nd ed.). Thousand Oaks, CA: Sage.

Sedikides, C., Meek, R., Alicke, M. D., & Taylor, S. 2014. Behind bars but above the bar: Prisoners consider themselves more prosocial than non-prisoners. *British Journal of Social Psychology*, 53: 396-403.

Shadish, W. R., Cook, T. D., & Campbell, D. T. 2002. *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston, MA: Houghton Mifflin.

Sheeran, P., & Webb, T. L. 2016. The intention-behavior gap. *Social and Personality Psychology Compass*, 10: 503-518.

Siemsen, E., Roth, A., & Oliveira, P. 2010. Common method bias in regression models with linear, quadratic, and interaction effects. *Organizational Research Methods*, 13: 456-476.

Siraz, S., Claes, B., De Castro, J. O., & Vaara, E. 2023. Theorizing the grey area between legitimacy and illegitimacy. *Journal of Management Studies*, 60: 942-962.

Suchman, M. C. 1995. Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20: 571-610.

Suddaby, R., Bitektine A., & Haack, P. 2017. Legitimacy. *Academy of Management Annals*, 11: 451-478.

Tost, L. P. 2011. An integrative model of legitimacy judgments. *Academy of Management Review*, 36: 686-710.

Tyler, T. R. 1997. The psychology of legitimacy: A relational perspective on voluntary deference to authorities. *Personality and Social Psychology Review*, 1: 323-345.

Vu, L., Soraperra, I., Leib, M., van der Weele, J., & Shalvi, S. 2023. Ignorance by choice: A meta-analytic review of the underlying motives of willful ignorance and its consequences. *Psychological Bulletin*, 149: 611-635.

Xiao, E., & Houser, D. 2005. Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences*, 102: 7398-7401.

Zhang, J., Deephouse, D. L., van Gorp, D., & Ebbers, H. 2022. Individuals' perceptions of the legitimacy of emerging market multinationals: Ethical foundations and construct validation. *Journal of Business Ethics*, 176: 801-825.

Zimmermann, F. 2020. The dynamics of motivated beliefs. *American Economic Review*, 110: 337-363.

FOOTNOTE

[1] Note that we did not cluster the standard errors at the group level as the groups are only used to calculate the final payoffs for the participants and the charity. There are no interactions between group members and Participants B are only exposed to Participant A's decision after they have made their own decisions.

TABLES

Table 1

Allocation Options and the Corresponding Payoffs for the Three Parties

Allocation option	Amount kept/taken by the decision-maker	Amount given to/kept by the charity	Bonus for half of the group
1	$0.00	$1.00	$0.00
2	$0.10	$0.90	$0.00
3	$0.20	$0.80	$0.00
4	$0.30	$0.70	$0.00
5	$0.40	$0.60	$0.00
6	$0.50	$0.50	$0.00
7	$0.60	$0.40	$0.05
8	$0.70	$0.30	$0.10
9	$0.80	$0.20	$0.15
10	$0.90	$0.10	$0.20
11	$1.00	$0.00	$0.25

Table 2

Comparison of Study 1 and Study 2

	Study 1	**Study 2**
Data collection platform	MTurk	Prolific
Propriety judgments manipulation	Yes	Yes
Self-interest manipulation	Yes	No
Norm violation manipulation	Yes	Yes
Participants A in a managerial position	No	Yes
Number of observations	1,889	801

Table 3

Average Treatment Effects on Propriety Judgments

	Propriety Judgments
Norm-violating framing	-0.104
	(0.137)
Self-Interest	0.032
	(0.133)
Norm-violating framing*Self-Interest	-0.15
	(0.194)
Favoring Participant A	0.163
	(0.113)
Norm-violating framing*Favoring Participant A	-0.345
	(0.177)
Self-Interest*Favoring Participant A	-0.08
	(0.168)
Norm-violating framing*Self-Interest*Favoring Participant A	0.095
	(0.255)
Option	-0.142***
	(0.009)
Constant	5.681***
	(0.103)
N	10241
R2	0.074

*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1

Table 4

Average Treatment Effects on Behavioral Support with and without Prior Measurement of Propriety Judgments

	Behavioral Support
Norm-violating framing	-0.272**
	(0.088)
Self-Interest	-0.244**
	(0.09)
Norm-violating framing*Self-Interest	0.136
	(0.128)
Favoring Participant A	-0.231**
	(0.085)
Norm-violating framing*Favoring Participant A	-0.086
	(0.123)
Self-Interest*Favoring Participant A	0.276*
	(0.126)
Norm-violating framing*Self-Interest*Favoring Participant A	0.046
	(0.181)
Prior Judgment	-0.288**
	(0.087)
Norm-violating framing*Prior Judgment	0.38**
	(0.123)
Self-Interest*Prior Judgment	0.353**
	(0.124)
Norm-violating framing*Self-Interest*Prior Judgment	-0.414*
	(0.177)
Favoring Participant A*Prior Judgment	0.248*
	(0.117)
Norm-violating framing*Favoring Participant A*Prior Judgment	-0.036
	(0.166)
Self-Interest*Favoring Participant A*Prior Judgment	-0.257
	(0.165)
Norm-violating framing*Self-Interest*Favoring Participant A*Prior Judgment	-0.095
	(0.243)
Option	-0.099***
	(0.004)
Constant	4.377***
	(0.064)
N	20779
R2	0.091

*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1

Table 5

Average Treatment Effects on Propriety Judgments in Study 2

	Propriety Judgments
Norm-violating framing	-0.195
	(0.143)
Favoring Participant A	-0.209†
	(0.119)
Norm-violating framing*Favoring Participant A	-0.153
	(0.196)
Option	-0.198***
	(0.015)
Constant	5.797***
	(0.122)
N	4532
R2	0.164

*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1

Table 6

Average Treatment Effects on Behavioral Support with and without Prior Measurement of Propriety Judgments in Study 2

	Behavioral Support
Norm-violating framing	-0.101
	(0.098)
Favoring Participant A	0.003
	(0.084)
Norm-violating framing*Favoring Participant A	-0.175
	(0.136)
Prior Judgment	0.060
	(0.091)
Norm-violating framing*Prior Judgment	-0.80
	(0.135)
Favoring Participant A*Prior Judgment	-0.204†
	(0.123)
Norm-violating framing*Favoring Participant A*Prior Judgment	0.127
	(0.186)
Option	-0.127***
	(0.007)
Constant	4.294***
	(0.073)
N	8811
R2	0.141
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1	

FIGURES

Figure 1

Moral Evaluation as a Three-Party Interaction (Adapted from DeScioli and Kurzban, 2009)

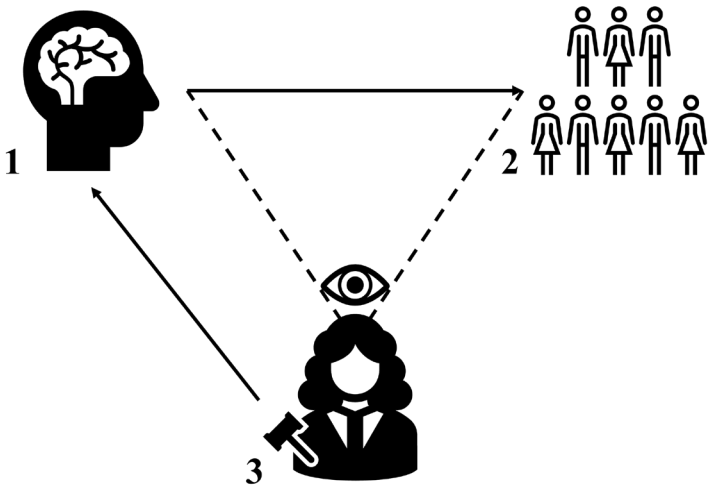


Figure 2

Overview of the Hypotheses

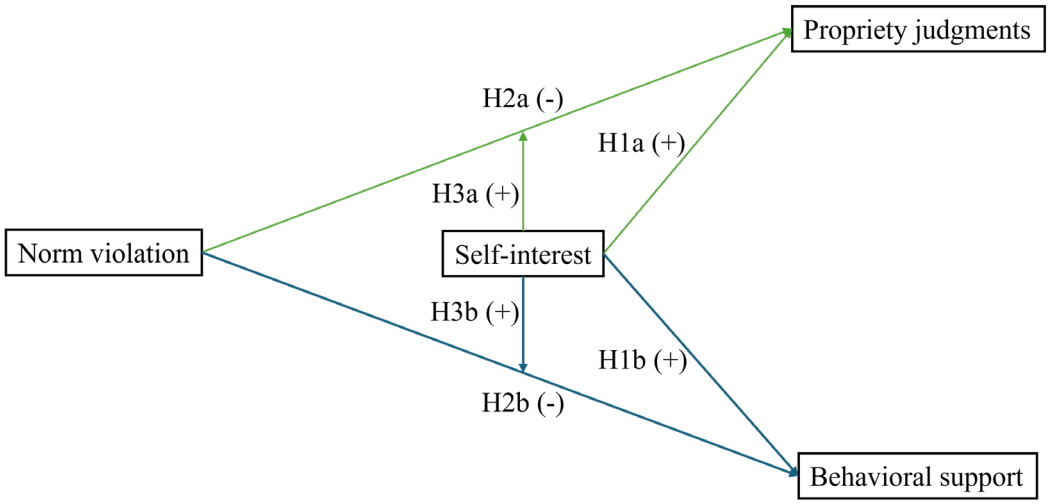


Figure 3

Experimental Design as a Three-Party Interaction

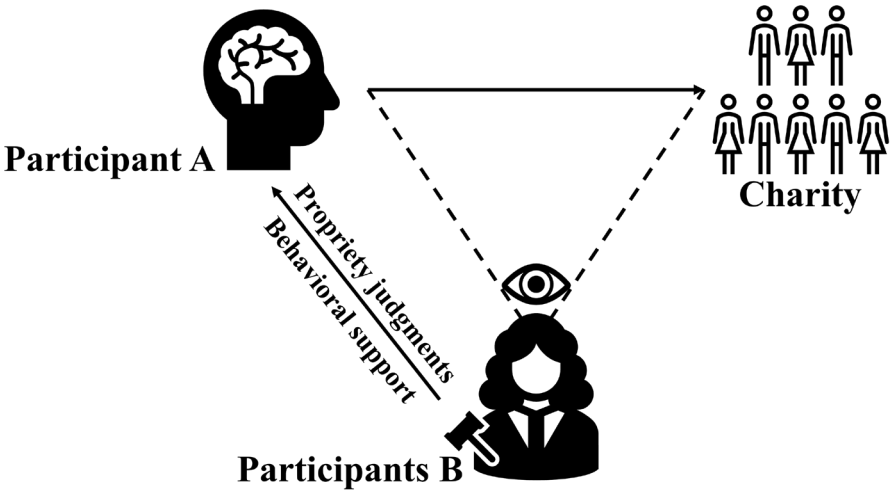


Figure 4

Average Propriety Judgments for Each Allocation Option Across Treatments, Including Baseline Comparisons

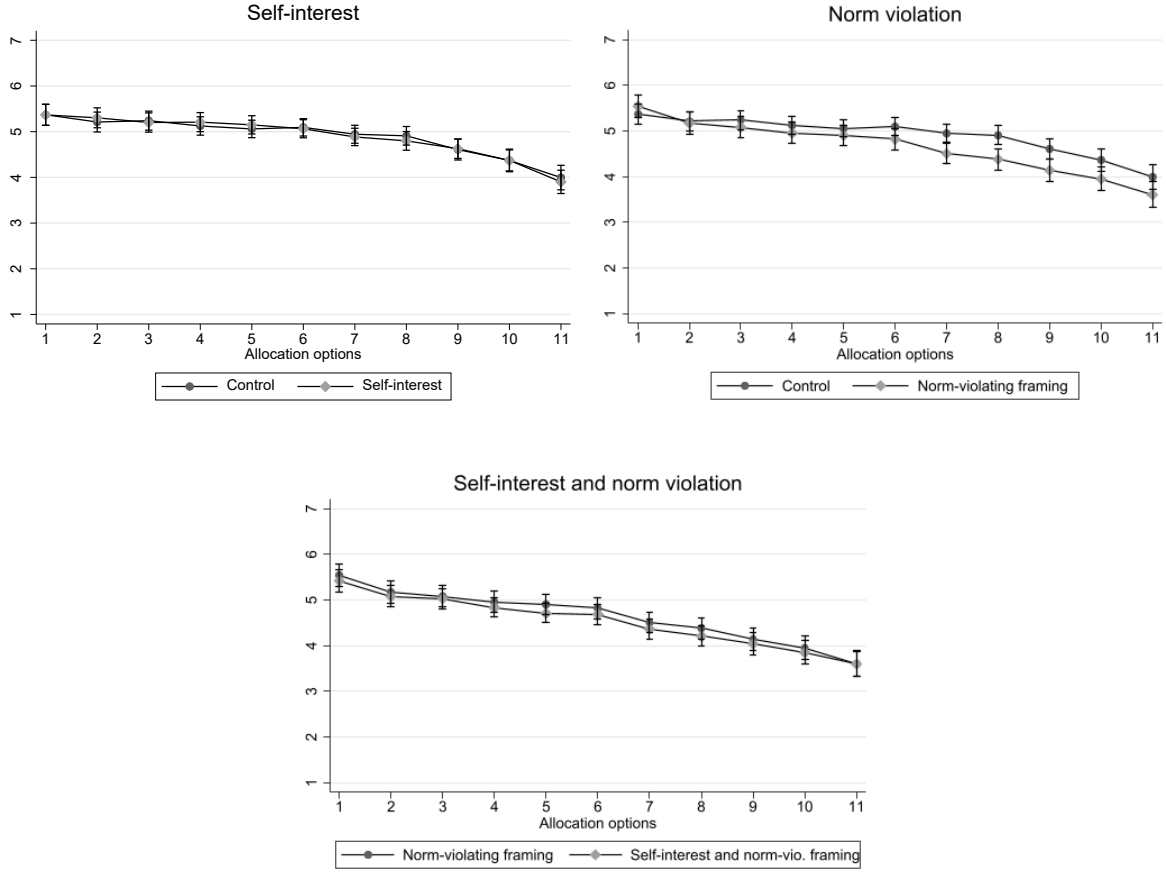


Figure 5

Average Behavioral Support without Prior Measurement of Propriety Judgments for Each Allocation Option Across Treatments, Including Baseline Comparisons

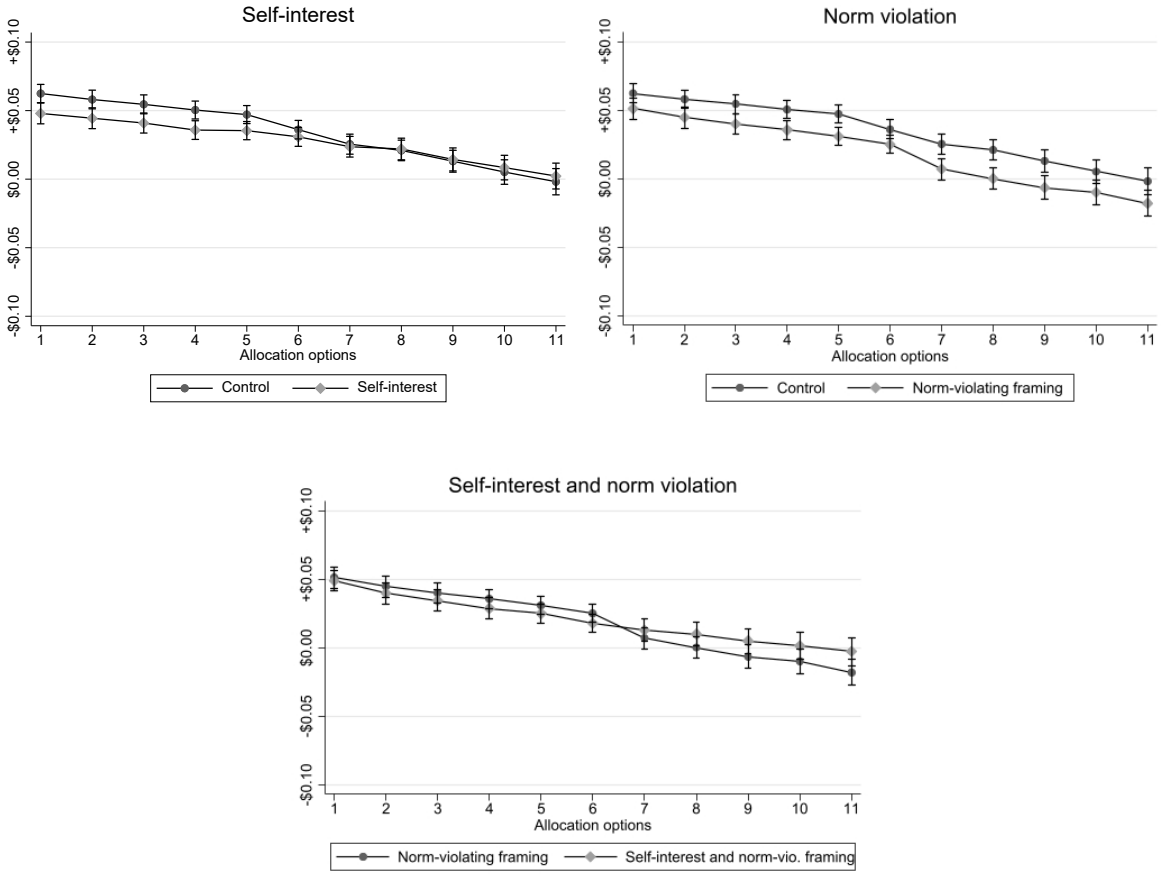


Figure 6

Average Treatment Effects on Behavioral Support with and without Prior Measurement of Propriety Judgments for Allocation Options Favoring the Charity and Participant A

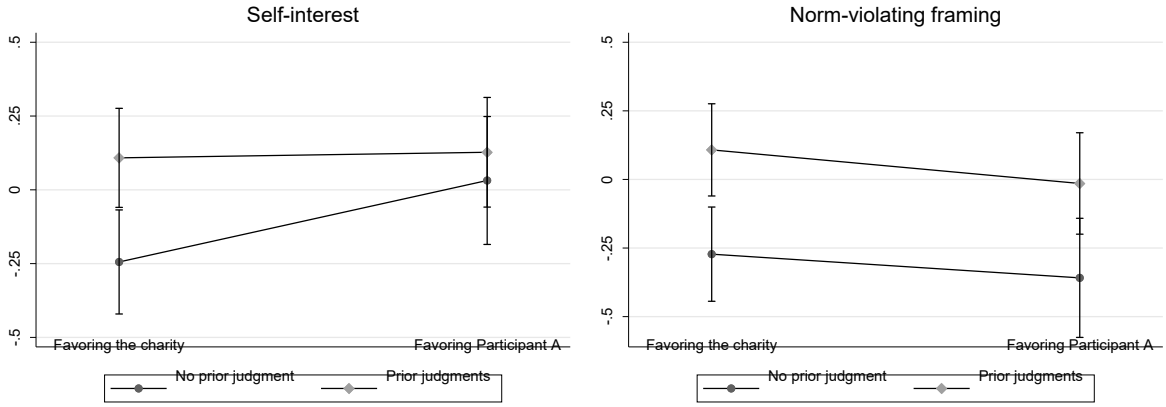


Figure 7

Average Propriety Judgments and Behavioral Support for Each Allocation Option Across Framings

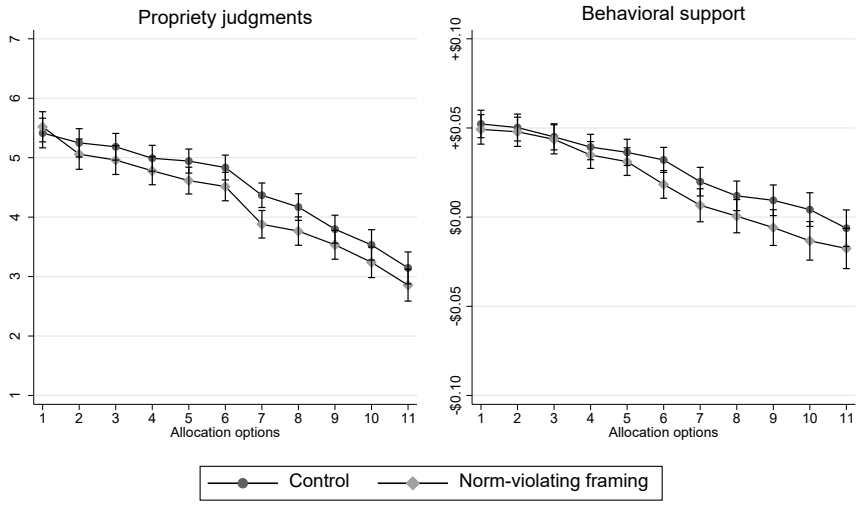
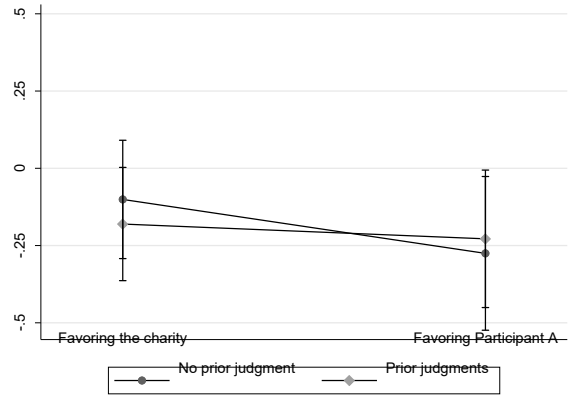


Figure 8

Average Treatment Effects on Behavioral Support with and without Prior Measurement of Propriety Judgments for Allocation Option Favoring the Charity and Participant A



ONLINE SUPPLEMENT

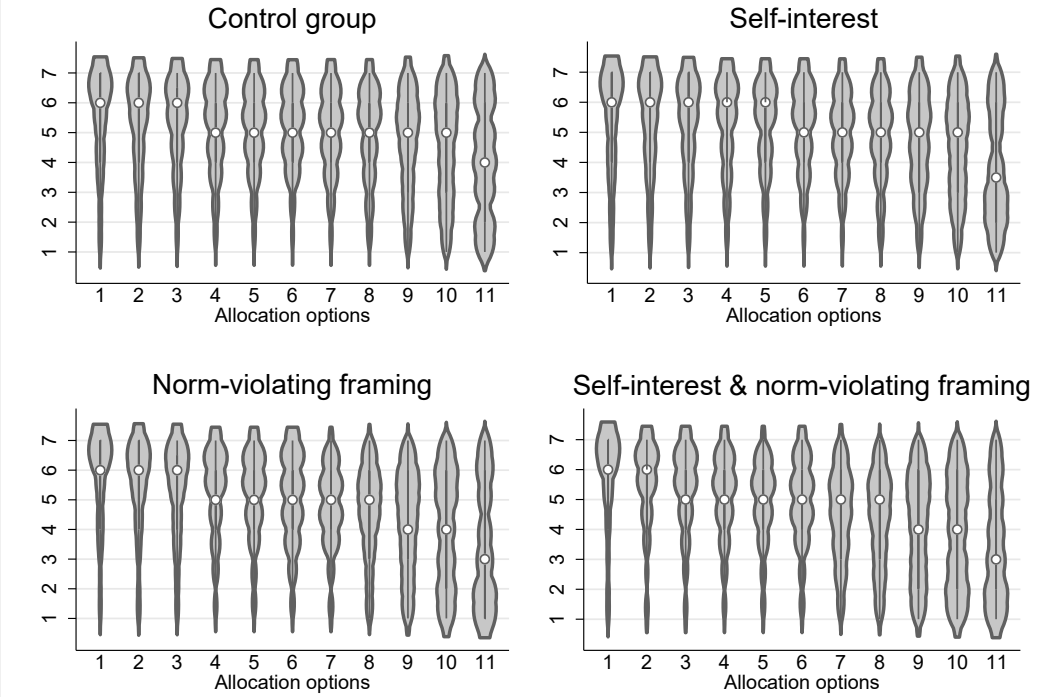
Appendix A

Experimental Instructions Provided to Participants

Propriety judgment	Control group	Self-interest	Control group	Norm-violating framing
Please evaluate the possible choices and indicate, for each of them, whether you personally believe choosing that option would be 'very inappropriate', 'inappropriate', 'somewhat inappropriate', 'neither inappropriate nor appropriate', 'somewhat appropriate', 'appropriate', or 'very appropriate'. In each of your responses, please answer as truthfully as possible, based on your personal belief of what constitutes appropriate or inappropriate behavior.	Half of the participants in your group will receive a bonus depending on the allocation chosen by the individual who participated in last week's HIT [human intelligence task]. The table below summarizes the possible allocations and their monetary consequences. You are not part of the participants in your group who will receive a bonus depending on the choice made by the individual who participated in last week's HIT.	Half of the participants in your group will receive a bonus depending on the allocation chosen by the individual who participated in last week's HIT [human intelligence task]. The table below summarizes the possible allocations and their monetary consequences. You are part of the participants in your group who will receive a bonus depending on the choice made by the individual who participated in last week's HIT.	The individual who participated in last week's HIT [human intelligence task] was allocated a bonus of $1.00 and was required to choose how much to give to the charity.	The charity was allocated a bonus of $1.00 and the individual who participated in last week's HIT [human intelligence task] was required to choose how much to take from the charity.

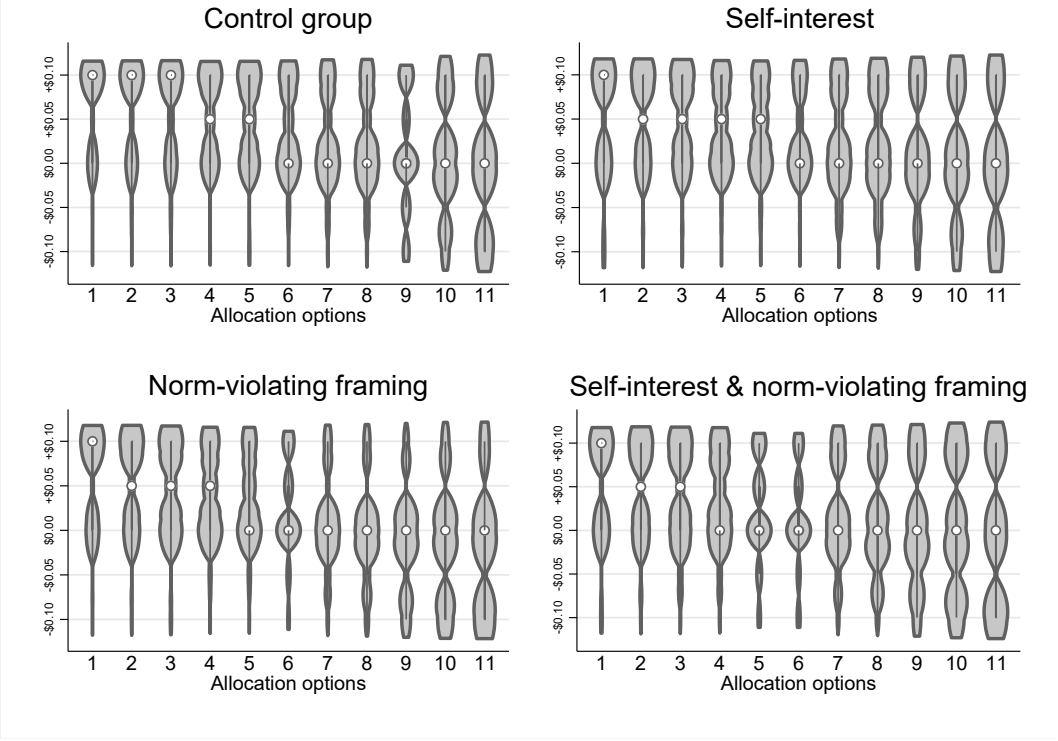
Appendix B

Kernel Distributions of the Propriety Judgments for Each Condition (Study 1)



Appendix C

Kernel Distributions of the Behavioral Support Without Prior Measurement of Propriety Judgments for Each Condition (Study 1)

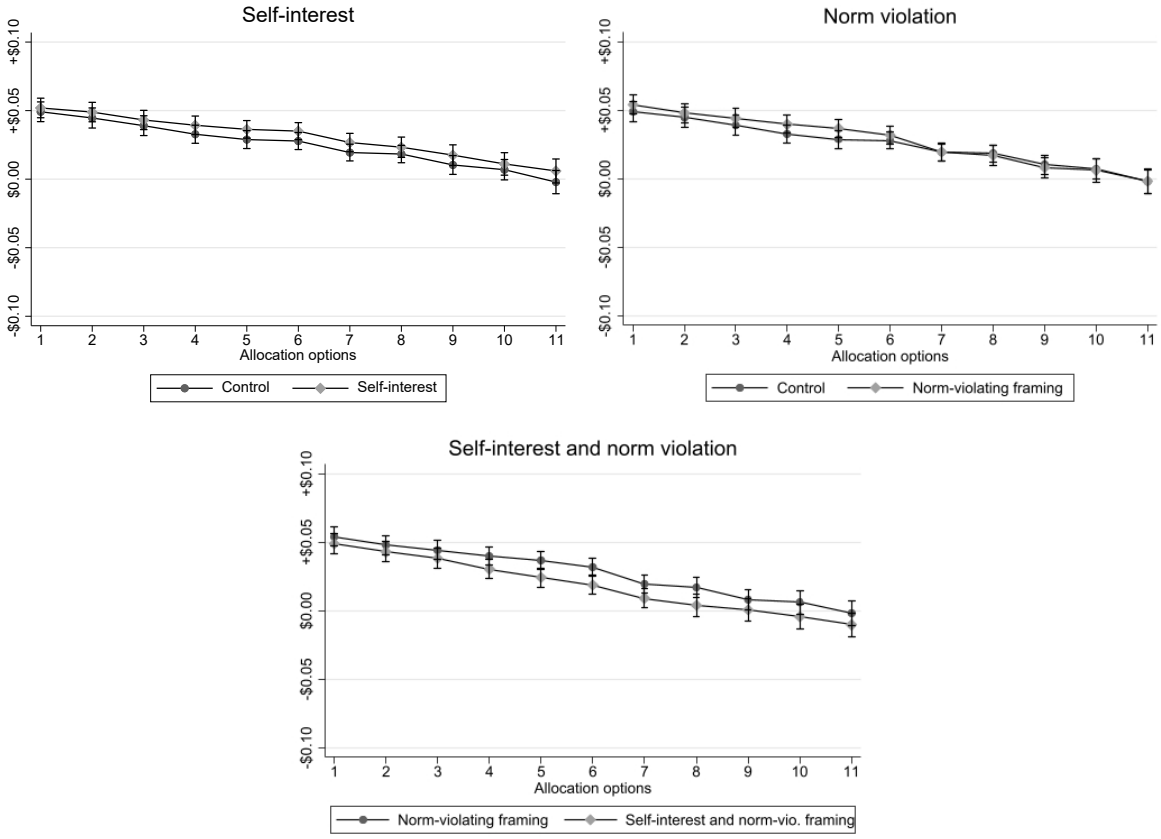


Appendix D

Kernel Distributions of the Behavioral Support with Prior Measurement of Propriety Judgments for each Condition (Study 1)

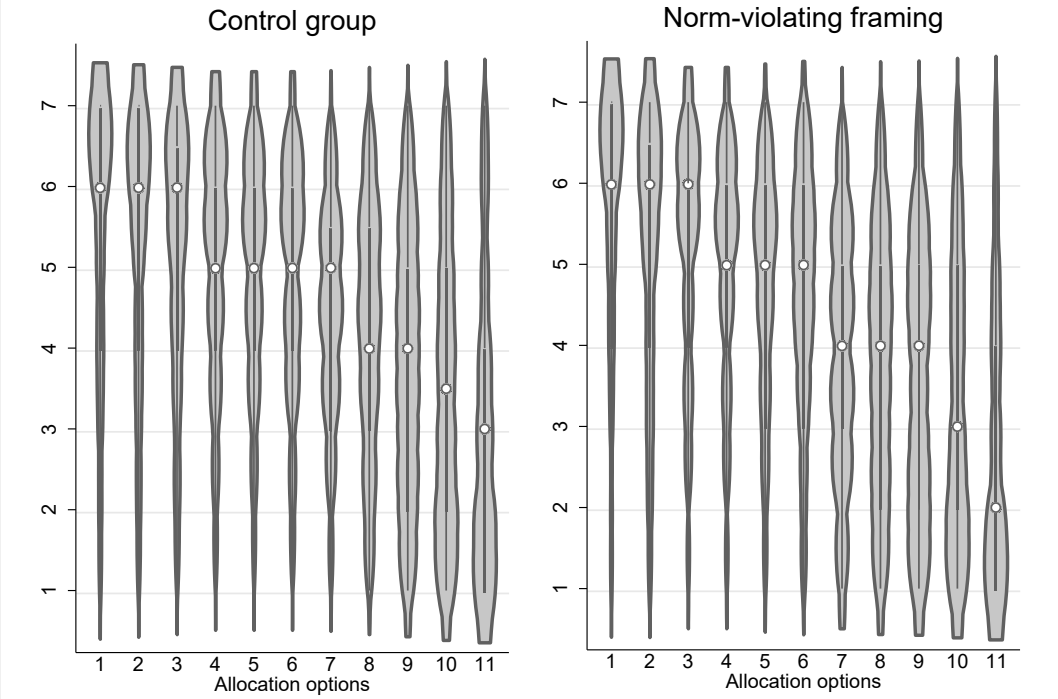
Appendix E

Average Behavioral Support with Prior Measurement of Propriety Judgments for Each Allocation Option Across Treatments, Including Baseline Comparisons (Study 1)



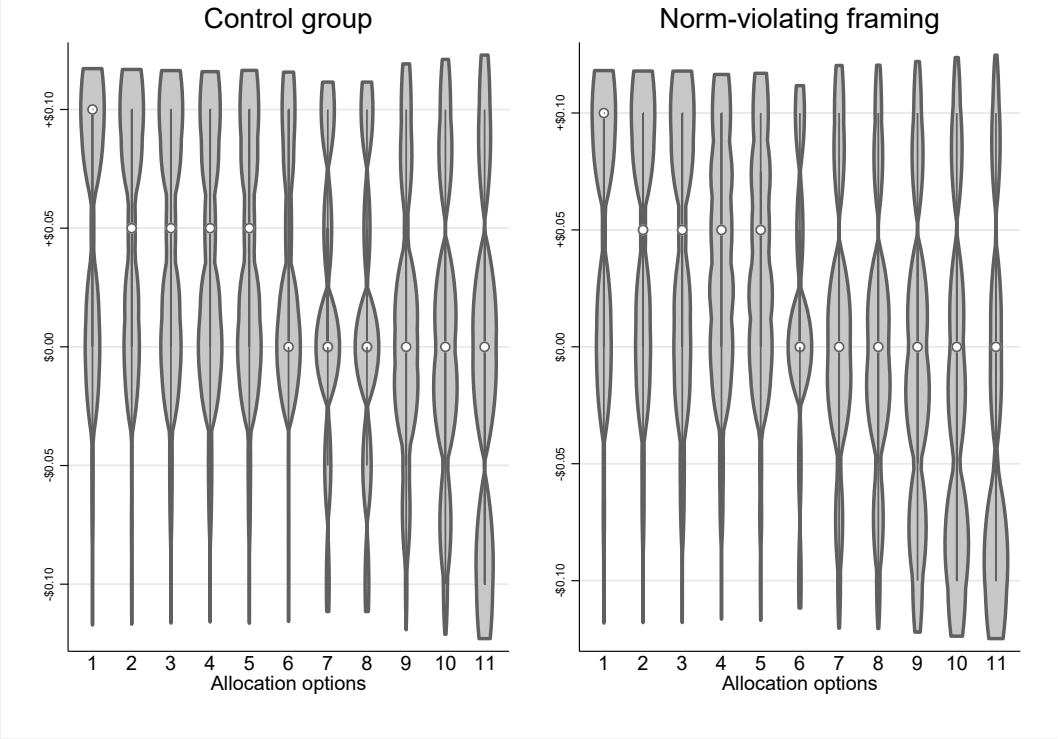
Appendix F

Kernel Distributions of the Propriety Judgments for Each Framing (Study 2)



Appendix G

Kernel Distributions of the Behavioral Support without Prior Measurement of Propriety Judgments for Each Framing (Study 2)



Appendix H

Kernel Distributions of the Behavioral Support with Prior Measurement of Propriety Judgments for Each Framing (Study 2)