

Article

Not peer-reviewed version

---

# Navigating the Paradox of Ethical Transgression: Managerial Robin Hoodism and the Boundaries of Supererogatory Justice Enactment

---

[Jonathan H. Westover](#)\*

Posted Date: 26 March 2026

doi: 10.20944/preprints202603.2089.v1

Keywords: managerial Robin Hoodism; supererogation; organizational justice; ethical decision-making; moral psychology; justice enactment



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Navigating the Paradox of Ethical Transgression: Managerial Robin Hoodism and the Boundaries of Supererogatory Justice Enactment

Jonathan H. Westover

Nexus Institute for Work & AI - Catalyst Center for Work Innovation; jon.westover@gmail.com

## Abstract

Managerial Robin Hoodism—the unauthorized redistribution of organizational resources to remedy perceived workplace injustice—occupies an ambiguous position in organizational ethics. This behavior simultaneously violates formal authority structures while pursuing morally laudable ends, creating classification difficulties for observers attempting to evaluate its moral status. Drawing on philosophical theories of supererogation and organizational justice enactment frameworks, we investigate how observers classify Robin Hoodism across the moral categories of obligatory, supererogatory, and prohibited action. Through three factorial vignette experiments with 423 working adults evaluating 847 systematically varied scenarios, we examine how situational characteristics and reflective processing shape these classifications. Study 1 establishes that observers most frequently classify Robin Hoodism as supererogatory (beyond duty) rather than obligatory or prohibited, though substantial variation exists across conditions. Study 2 demonstrates that injustice severity and victim marginalization shift classifications toward obligation, with moral outrage partially mediating these effects. Study 3 reveals that structured moral reflection significantly increases obligatory classifications at the expense of supererogatory judgments, suggesting that apparent optionality may reflect organizational framing rather than genuine moral discretion. We introduce the distinction between *thin supererogation*—apparent optionality that tends to shift toward perceived obligation under reflection—and *thick supererogation*—perceived optionality that remains stable under extended moral consideration. This distinction offers a conceptual vocabulary for understanding how organizational contexts may shape moral perception. It is important to emphasize that this distinction describes patterns in observer judgments rather than adjudicating the genuine moral status of actions—a point we develop throughout. Theoretical implications for justice enactment research and practical implications for organizational ethics are discussed, along with important limitations regarding the descriptive nature of these findings.

**Keywords:** managerial Robin Hoodism; supererogation; organizational justice; ethical decision-making; moral psychology; justice enactment

---

Organizations regularly produce injustices. Despite formal grievance mechanisms, equity policies, and ethical codes, employees continue to experience discrimination, procedural unfairness, and interpersonal mistreatment that official channels fail to remedy (Colquitt et al., 2013). When confronted with such failures, some managers take matters into their own hands—redistributing resources, adjusting schedules, or modifying performance evaluations to counteract perceived wrongs, all without formal authorization. This phenomenon, which we term *managerial Robin Hoodism*, presents a fundamental puzzle for organizational ethics: How should we classify actions that pursue justice through formally illegitimate means?

The ethical complexity of Robin Hoodism derives from its dual nature. On one hand, these behaviors violate organizational authority structures, circumvent established procedures, and potentially expose managers to significant professional risk. On the other hand, they represent attempts to remedy genuine injustice when formal mechanisms prove inadequate. This tension places

Robin Hoodism at the intersection of two bodies of scholarship that have developed largely in isolation: organizational justice research examining how individuals respond to workplace fairness violations, and moral philosophy investigating the boundaries between obligatory and supererogatory action.

The question of whether Robin Hoodism constitutes supererogatory action—morally praiseworthy but optional behavior that exceeds duty's demands—carries significant implications for how we understand managerial moral agency. If such behavior is genuinely supererogatory, managers who refrain from unauthorized justice interventions commit no wrong; they simply decline to perform heroic acts beyond what morality requires. However, if apparent supererogation masks genuine obligation—if managers *ought* to intervene when formal channels fail—then framing Robin Hoodism as optional may provide unwarranted moral cover for inaction in the face of injustice.

This research investigates how observers classify Robin Hoodism across the moral categories of obligatory, supererogatory, and prohibited action, and examines the conditions under which these classifications shift. We pursue three primary objectives. First, we establish baseline classification patterns, examining which moral category observers most frequently apply to Robin Hoodism and how robust these classifications are across varying conditions. Second, we identify situational factors that influence classification, testing whether characteristics of the injustice, the victim, and the availability of alternatives shift judgments toward obligation or toward optionality. Third, we examine whether structured moral reflection alters classification patterns, testing whether supererogatory judgments remain stable or shift toward obligation when observers engage in systematic moral reasoning.

Our findings contribute to organizational scholarship in several ways. Empirically, we document substantial variation in how observers classify identical behaviors depending on situational characteristics and reflective processing. Theoretically, we introduce the distinction between *thin supererogation*—apparent optionality that tends to dissolve under reflection—and *thick supererogation*—perceived optionality that remains stable under extended moral consideration. This distinction provides a conceptual vocabulary for understanding how organizational contexts may shape moral perception. Practically, our findings suggest that organizations may benefit from creating conditions that support rather than penalize justice-motivated interventions, while respecting the genuine moral diversity that exists among employees regarding the permissibility of such actions.

We proceed as follows. First, we review literature on unauthorized workplace justice interventions and philosophical supererogation, identifying the theoretical puzzle motivating our research. Second, we develop hypotheses regarding classification patterns, situational moderators, and reflection effects. Third, we present three factorial vignette experiments testing these hypotheses. Fourth, we discuss theoretical contributions, practical implications, and important limitations. We conclude by outlining directions for future research on the boundaries of moral obligation in organizational contexts.

## Theoretical Foundations

### *Unauthorized Justice Interventions in Organizations*

The Robin Hood legend endures across cultures because it crystallizes a fundamental moral tension: the conflict between formal authority and substantive justice. When legitimate channels fail to remedy injustice, individuals face a choice between accepting unjust outcomes and pursuing justice through unauthorized means. In organizational contexts, this tension manifests when managers use their discretionary authority in unauthorized ways to redistribute resources, opportunities, or evaluations in an attempt to correct or remedy perceived organizational injustices.

We use the term *managerial Robin Hoodism* to describe this phenomenon—the use of discretionary managerial behaviors to redistribute organizational resources in ways intended to remedy perceived injustice, enacted outside or against formal organizational authority structures. This construct shares conceptual space with related phenomena in the organizational literature, including prosocial rule

breaking (Morrison, 2006), constructive deviance (Galperin, 2012), and counter-role behaviors undertaken to benefit others at potential cost to oneself.

Morrison (2006) introduced the concept of prosocial rule breaking to describe instances where employees violate organizational rules with the intention of benefiting the organization or its stakeholders. This work established that rule-breaking is not uniformly antisocial but can serve constructive purposes. Galperin (2012) extended this line of inquiry by examining constructive deviance—behaviors that depart from organizational norms but are intended to benefit the organization. Both constructs recognize that formal organizational rules may sometimes conflict with organizational or stakeholder welfare, and that employees may navigate this tension by prioritizing welfare over compliance.

Robin Hoodism represents a specific form of prosocial rule breaking characterized by its explicitly justice-oriented motivation. While prosocial rule breaking encompasses diverse behaviors undertaken for various beneficial purposes (efficiency, customer service, colleague assistance), Robin Hoodism specifically involves redistribution intended to remedy unfairness. This justice orientation connects Robin Hoodism to literature on organizational justice and positions it as a form of justice enactment—active efforts to create, maintain, or restore fairness (Scott et al., 2014).

Research on employee responses to injustice has documented various reactions, including reduced organizational commitment, withdrawal behaviors, and retaliatory actions (Ambrose & Schminke, 2009). Less attention has focused on constructive responses that attempt to remedy injustice directly. The prosocial rule breaking literature suggests that such responses occur, but the specific phenomenon of justice-motivated resource redistribution by managers has received limited empirical investigation.

What distinguishes Robin Hoodism from ordinary managerial discretion is its unauthorized character combined with its remedial intent. Managers routinely exercise discretion in allocating assignments, providing feedback, and distributing opportunities. Such discretion becomes Robin Hoodism when it is exercised outside or against formal authority structures specifically to counteract perceived injustice—for example, providing favorable performance ratings to compensate for discriminatory treatment, or reallocating desirable assignments to remedy unfair distribution by others.

### *The Philosophy of Supererogation*

Philosophical analysis of supererogation—literally, paying beyond what is owed—provides conceptual resources for examining questions about the moral status of Robin Hoodism. The category of supererogation encompasses actions that are morally good but not morally required, such that omitting them constitutes no wrong (Heyd, 2019). Classic examples include heroic rescue at significant personal risk, extraordinary charitable giving, and forgiveness beyond what justice demands. Such actions merit praise when performed but warrant no blame when omitted.

The concept's theoretical foundations remain contested. Consequentialist frameworks struggle to accommodate supererogation because they imply that agents always ought to produce the most good possible, leaving no space for praiseworthy actions beyond duty (Horgan & Timmons, 2010). If maximizing good is obligatory, actions producing more good than alternatives are required rather than optional. Some consequentialists respond by incorporating agent-relative permissions, allowing agents to give extra weight to their own projects and relationships, thereby creating space for actions that exceed the resulting reduced demands (Scheffler, 1982).

Deontological frameworks accommodate supererogation more naturally by distinguishing between *perfect duties*—specific obligations to or against particular actions—and *imperfect duties*—general requirements admitting latitude in execution (Kant, 1797/1996). Imperfect duties to beneficence, for instance, require that agents sometimes help others but leave discretion regarding when, whom, and how much to help. Actions exceeding minimal compliance with imperfect duties may be supererogatory in the sense that agents go beyond what duty specifically requires while fulfilling the general obligation.

Virtue ethics approaches supererogation differently, emphasizing character development over action classification. From this perspective, supererogatory actions may both express and cultivate virtues such as compassion, courage, and justice (Crisp, 2013). The virtuous person is disposed to act well even when doing so is not required, and supererogatory actions provide occasions for virtue expression and development. This framing suggests that even if Robin Hoodism is not obligatory, it may be important for managerial character formation.

A critical distinction for our purposes concerns the difference between *qualified* and *unqualified* supererogation. As Heyd (1982, 2019) develops the point, unqualified supererogation involves actions that genuinely exceed duty—they would not be required under any reasonable interpretation of the agent's obligations. Qualified supererogation involves actions that may nominally exceed formal role requirements but respond to moral considerations that, upon reflection, might generate genuine obligations. This distinction anticipates our conceptual contribution: the differentiation between thick and thin supererogation.

#### *Organizational Contexts and Moral Perception*

A key question arising from philosophical analysis concerns whether organizational contexts influence how observers perceive moral requirements. Several theoretical traditions suggest they should. Role-based theories of moral obligation hold that organizational positions generate specific duties attached to roles rather than persons (Simmons, 1979). Managers may have role-specific duties that exceed general moral requirements—fiduciary obligations to employees, duties of fair treatment attached to supervisory positions, and responsibilities for organizational justice within their domains. These role-based duties could transform what would be supererogatory for ordinary individuals into obligations for those occupying specific positions.

Institutional theories emphasize how organizations shape moral perception through taken-for-granted assumptions about appropriate behavior (Zucker, 1987). What appears optional may reflect institutionalized understandings rather than genuine moral discretion. Organizations routinely frame matters of potential obligation as matters of discretion—through language of “going above and beyond,” “organizational citizenship,” and “discretionary effort”—in ways that may obscure underlying moral requirements. This framing typically serves organizational interests by minimizing claims employees might make on organizational resources while praising exceptional contributions.

Critical management perspectives extend this analysis, suggesting that organizational discourse may systematically shape perceptions of power relations and moral obligations (Alvesson & Willmott, 1992). From this perspective, the framing of Robin Hoodism as supererogatory rather than obligatory may serve organizational interests by placing remedial burdens on individual managers while reducing organizational responsibility for systemic justice failures.

These theoretical traditions converge on a hypothesis: that supererogatory classifications of Robin Hoodism may reflect organizational framing rather than settled moral judgment. If so, we would expect such classifications to be unstable—vulnerable to revision when observers step back from organizational perspectives and engage in systematic moral reflection. This hypothesis motivates our examination of how reflection affects moral classification.

#### *Justice Enactment and Moral Obligation*

Recent organizational justice scholarship has moved beyond studying reactions to justice and injustice toward examining how individuals actively create justice through their behaviors—what researchers term *justice enactment* (Scott et al., 2014). This shift reframes organizational members as moral agents capable of shaping fairness conditions rather than merely responding to them.

Justice enactment research demonstrates that individuals routinely engage in behaviors intended to create, maintain, or restore fairness. Managers enact justice through consistent procedure application, transparent communication, and respectful treatment; employees enact justice through peer support, information sharing, and informal sanctioning of unfair behavior (Sherf et al., 2019).

Robin Hoodism represents an extreme form of justice enactment—one that pursues fairness through formally unauthorized means.

A key question concerns when justice enactment represents moral obligation versus moral virtue. The philosophical literature on duties of rescue provides relevant analysis. The duty to rescue those in peril—paradigmatically, pulling a drowning child from a shallow pond—is widely considered obligatory when rescue is possible at minimal cost (Singer, 1972). Debate concerns how demanding rescue duties are: Must agents sacrifice significantly to rescue others, or only provide easy assistance? The literature suggests that obligation strength increases with victim need, agent capacity, and absence of alternatives, while decreasing with cost to the agent (Scanlon, 1998).

Applied to Robin Hoodism, this analysis suggests that justice enactment might move from supererogatory to obligatory as injustice becomes severe, victims become vulnerable, alternatives become unavailable, and intervention costs remain manageable. We develop these ideas into testable hypotheses in the following section.

## Hypothesis Development

### *Baseline Classifications*

The first empirical question concerns how observers initially classify Robin Hoodism. Theoretical considerations support competing predictions. On one hand, philosophical analysis suggests that Robin Hoodism involves genuine moral complexity—combining praiseworthy ends with procedurally problematic means—that might lead observers toward supererogatory classifications that recognize moral value without implying obligation. The procedural violations inherent in Robin Hoodism distinguish it from ordinary benevolence and may create sufficient moral ambiguity to render the behavior optional rather than required.

On the other hand, organizational contexts may frame behaviors that respond to genuine moral requirements as discretionary. If Robin Hoodism addresses real injustice that organizations have failed to remedy, the language of “going above and beyond” may obscure underlying obligations rather than reflect them. The question is empirical: How do observers actually classify these behaviors?

We expect that observers will most frequently classify Robin Hoodism as supererogatory, for several reasons. First, the organizational framing of managerial intervention as discretionary is deeply institutionalized and likely to shape initial perception. Second, the procedural violations inherent in Robin Hoodism create moral complexity that observers may resolve by withholding the strong judgment of obligation. Third, supererogatory classification allows observers to praise the behavior without condemning those who refrain. We therefore hypothesize:

**Hypothesis 1.** *Observers will classify Robin Hoodism as supererogatory more frequently than as obligatory or prohibited.*

This hypothesis concerns initial classifications. We remain agnostic about whether such classifications reflect considered moral judgment or institutionalized framing, addressing this question in subsequent hypotheses.

### *Situational Moderators*

Even if supererogatory classifications predominate overall, we expect situational characteristics to moderate classification patterns. Three factors derived from philosophical analysis of rescue duties seem particularly relevant: injustice severity, victim marginalization, and alternative availability.

*Injustice severity* should affect classification because more severe injustice generates stronger claims on potential rescuers. As harm increases, the moral stakes of intervention rise, and what might be optional for minor injustice may become obligatory for severe injustice. This reasoning aligns with philosophical analysis of scalar properties of obligation: duty strength varies with morally relevant features rather than applying uniformly across cases (Brink, 1994). We hypothesize:

**Hypothesis 2a.** *Observers will more frequently classify Robin Hoodism as obligatory when injustice severity is high (versus low).*

*Victim marginalization* should affect classification because marginalized individuals face systematic disadvantages that amplify injustice's impact and limit self-remedy capacity. Historical discrimination, resource deprivation, and social exclusion may generate stronger claims on those with power to help (Young, 2011). A wrong against someone already disadvantaged may warrant intervention that would be optional for someone with greater resources to pursue remedy independently. We hypothesize:

**Hypothesis 2b.** *Observers will more frequently classify Robin Hoodism as obligatory when the victim is marginalized (versus non-marginalized).*

*Alternative availability* should affect classification because obligation arises more clearly when the agent represents the only remedy source. If formal channels remain available and potentially effective, Robin Hoodism may represent one option among many, reducing any particular intervention's obligatory character. When alternatives have been exhausted without success, however, the manager may be the victim's last resort, strengthening obligatory classification. This reasoning aligns with philosophical analysis of "last resort" conditions in rescue duty (Miller, 2004). We hypothesize:

**Hypothesis 2c.** *Observers will more frequently classify Robin Hoodism as obligatory when alternatives have been exhausted (versus when alternatives remain available).*

#### *The Mediating Role of Moral Outrage*

We expect that the effects of situational characteristics on classification operate partly through emotional response, specifically moral outrage. Moral outrage—anger at violations of moral standards—represents an emotional appraisal that injustice has occurred and response is warranted (Batson et al., 2007). Unlike personal anger motivated by self-interest, moral outrage responds to violations affecting others and motivates action even at personal cost.

Moral outrage may function as a mechanism linking situational characteristics to obligatory classification. Severe injustice and victim marginalization should evoke greater outrage, which may in turn heighten perceived obligation to respond. Emotions are not merely reactions to cognitive judgments; they can constitute moral appraisals that inform subsequent judgment (Haidt, 2001). Outrage may signal—accurately or not—that the situation demands response, potentially shifting classification toward obligation.

We note that the role of emotion in moral judgment is contested. Some philosophers argue that emotions like outrage provide genuine moral knowledge (Tappolet, 2016), while others contend that emotional responses may bias judgment and should be overridden by cooler reflection (Kahneman, 2011). Our mediation analysis tests whether outrage statistically accounts for relationships between situational factors and classification, remaining agnostic about whether such mediation reflects accurate moral tracking or emotional bias. We hypothesize:

**Hypothesis 3.** *Moral outrage will mediate the relationships between (a) injustice severity, (b) victim marginalization, and obligatory classification.*

#### *The Effect of Structured Moral Reflection*

Our final hypothesis concerns whether classification patterns shift under structured moral reflection. If supererogatory classifications reflect settled moral judgment, they should remain stable when observers engage in systematic moral reasoning. However, if supererogatory classifications

reflect organizational framing that obscures underlying moral requirements, reflection may prompt reconsideration.

The hypothesis that reflection increases obligatory classification rests on several considerations. First, reflection may help observers recognize role-based duties attached to managerial positions that initial judgment overlooks. Second, reflection may prompt consideration of the manager's unique position—their awareness of injustice, capacity to intervene, and relationship to the victim—that generates agent-specific obligations. Third, reflection may surface the moral arbitrariness of organizational authority that initial classification takes for granted.

However, alternative outcomes are equally plausible. Reflection might *decrease* obligatory classification by surfacing considerations that support optionality—legitimate organizational interests in maintaining authority structures, potential negative consequences of unauthorized intervention, or respect for victim autonomy that counsels against paternalistic rescue. Reflection might also increase prohibited classification by highlighting the rule-violations inherent in Robin Hoodism.

We tentatively predict that reflection will increase obligatory classification, based on the theoretical argument that organizational framing may systematically obscure obligation. This hypothesis is more exploratory than confirmatory:

**Hypothesis 4.** *Structured moral reflection will increase the frequency of obligatory classifications and decrease the frequency of supererogatory classifications.*

## Overview of Studies

We tested these hypotheses across three factorial vignette experiments. All studies received institutional ethics approval and obtained informed consent from participants. Scenario materials, analysis code, and additional results are available in supplementary materials.

Factorial vignette methodology involves presenting participants with detailed scenarios that systematically vary factors of theoretical interest (Aguinis & Bradley, 2014). This approach combines the internal validity of experimental manipulation with the external validity of realistic narrative scenarios. Participants read scenarios describing managerial Robin Hoodism and classify the behavior into moral categories while reporting emotional responses and engaging in manipulation checks.

*Study 1* establishes baseline classification patterns, testing Hypothesis 1 regarding the predominance of supererogatory classifications. *Study 2* examines situational moderators, testing Hypotheses 2a-c and the mediation hypothesis (Hypothesis 3). *Study 3* tests whether structured moral reflection shifts classifications, examining Hypothesis 4 through a pre-post reflection design.

### Study 1: Baseline Classification Patterns

#### Method

##### Participants

We recruited 142 working adults through Prolific Academic, a platform providing high-quality participant samples for academic research (Palan & Schitter, 2018). Inclusion criteria required participants to be employed at least part-time, aged 25 or older, and located in the United States. These criteria ensured participants had workplace experience relevant to evaluating managerial scenarios.

Participants received 2.50 for approximately 12 minutes of participation (effective hourly rate: 12.50). The sample was 52% female, with a mean age of 38.4 years ( $SD = 10.2$ ) and mean organizational tenure of 7.3 years ( $SD = 6.1$ ). Regarding education, 41% held bachelor's degrees, 23% held graduate degrees, and 36% had completed some college or held associate degrees. Participants were drawn from diverse industries, including healthcare (18%), education (15%), technology (14%), retail (12%), and financial services (11%), with the remainder distributed across other sectors.

##### Materials and Procedure

We developed 12 vignettes depicting managerial Robin Hoodism across three injustice types (discriminatory treatment, procedural unfairness, and interpersonal mistreatment) crossed with two severity levels (moderate and severe) and two organizational contexts (for-profit corporation and non-profit organization). The following example illustrates the vignette structure:

*Jennifer is a mid-level manager at a [for-profit technology company / non-profit social services organization]. One of her team members, Marcus, was recently passed over for promotion despite having the strongest performance record in the department. Jennifer learned through a colleague that the promotion decision was heavily influenced by the senior director's personal bias against Marcus because of his [race/religious background]. Jennifer raised the issue with HR, but after a cursory investigation, they concluded there was "insufficient evidence" to take action.*

*Frustrated by this outcome, Jennifer decided to take matters into her own hands. She used her discretionary authority to give Marcus the most desirable project assignments, arranged for him to present at an important client meeting that would increase his visibility, and provided a more favorable performance rating than strictly warranted by formal metrics—all intended to compensate for the unfair promotion decision and position Marcus better for future advancement.*

Vignettes were developed through iterative piloting with 30 participants not included in the main study. Piloting assessed comprehension, realism, and clarity, with vignettes revised based on feedback. All final vignettes were rated as realistic (mean rating above 5 on a 7-point scale) and comprehensible (fewer than 5% of pilot participants indicating confusion).

Each participant read and evaluated four randomly selected vignettes, yielding 568 observations (142 participants × 4 scenarios). After each vignette, participants classified the manager's behavior using the following item:

*How would you classify the manager's actions?*

- *Obligatory:* The manager was morally required to act this way; failing to do so would have been wrong.
- *Supererogatory:* The manager's actions were morally good but not required; they went above and beyond what duty demands.
- *Prohibited:* The manager's actions were wrong; the manager should not have acted this way regardless of intentions.

Following classification, participants completed manipulation checks and demographic measures. Manipulation checks assessed perceived injustice severity (3 items,  $\alpha = .87$ ) and victim vulnerability (2 items,  $\alpha = .79$ ), confirming that the manipulations operated as intended.

#### *Analysis Strategy*

We analyzed classification frequencies using multilevel multinomial logistic regression with observations nested within participants. This approach accounts for non-independence of multiple responses from the same participant while testing whether classification proportions differ from chance. We report odds ratios comparing classification category frequencies and 95% confidence intervals.

#### *Results*

Table 1 presents classification frequencies across conditions. Overall, 47.2% of observations received supererogatory classifications, 32.4% received obligatory classifications, and 20.4% received prohibited classifications. The pattern differed significantly from equal distribution,  $\chi^2(2) = 73.41, p < .001$ .

**Table 1.** Classification Frequencies by Condition (Study 1).

Condition	Obligatory	Supererogatory	Prohibited	<i>n</i>
Overall	32.4%	47.2%	20.4%	568
Moderate severity	26.1%	50.8%	23.1%	284

Severe	38.7%	43.6%	17.7%	284
For-profit context	31.2%	46.1%	22.7%	286
Non-profit context	33.6%	48.3%	18.1%	282

Multinomial regression confirmed that supererogatory classifications were significantly more frequent than obligatory classifications (OR = 1.85, 95% CI [1.47, 2.33],  $p < .001$ ) and prohibited classifications (OR = 3.47, 95% CI [2.63, 4.58],  $p < .001$ ). These results support Hypothesis 1.

Severity affected classification patterns such that severe conditions increased obligatory classifications relative to moderate conditions (OR = 1.78, 95% CI [1.32, 2.40],  $p < .001$ ), providing preliminary support for Hypothesis 2a. Organizational context (for-profit versus non-profit) did not significantly affect classification, OR = 1.09, 95% CI [0.81, 1.47],  $p = .57$ .

Analysis by injustice type revealed modest differences in classification patterns. Discrimination scenarios yielded the highest obligatory classifications (36.8%), followed by procedural unfairness (32.1%) and interpersonal mistreatment (28.3%), though these differences did not reach statistical significance in pairwise comparisons after correction for multiple testing (all adjusted  $ps > .08$ ).

#### *Discussion*

Study 1 establishes that observers most frequently classify Robin Hoodism as supererogatory, supporting Hypothesis 1. Nearly half of all classifications fell into this category, compared to roughly one-third obligatory and one-fifth prohibited. This pattern suggests that organizational members predominantly view justice-motivated rule-breaking as praiseworthy but optional rather than morally required.

However, the substantial presence of both obligatory and prohibited classifications indicates meaningful disagreement. One-third of observations were judged obligatory, suggesting that many observers view inaction in the face of injustice as morally problematic. One-fifth were judged prohibited, indicating that significant minorities view unauthorized intervention as impermissible regardless of motivation. This moral diversity warrants attention that we return to in the General Discussion.

Preliminary evidence also supports the moderating role of severity: observers classified Robin Hoodism as obligatory more frequently when the underlying injustice was severe. Study 2 provides a more rigorous test of situational moderators, including severity, marginalization, and alternative availability, and examines moral outrage as a mediating mechanism.

### **Study 2: Situational Moderators and Mediation**

#### **Method**

##### *Participants*

We recruited 156 working adults through Prolific Academic using the same inclusion criteria as Study 1. Participants received \$3.00 for approximately 15 minutes of participation. The sample was 49% female, with a mean age of 39.7 years ( $SD = 11.1$ ) and mean organizational tenure of 8.1 years ( $SD = 6.8$ ). Demographic composition was similar to Study 1 across education and industry.

##### *Design*

Study 2 employed a 2 (injustice severity: moderate vs. severe)  $\times$  2 (victim marginalization: marginalized vs. non-marginalized)  $\times$  2 (alternative availability: exhausted vs. available) between-subjects factorial design. Participants were randomly assigned to one of eight conditions and evaluated a single detailed scenario, yielding 156 observations with approximately 19-20 per cell.

##### *Materials*

Scenarios described a managerial response to workplace discrimination. The base scenario described a situation where an employee was unfairly denied a training opportunity. Experimental factors were manipulated as follows:

*Injustice severity.* In the moderate condition, the employee missed one desirable training opportunity. In the severe condition, the employee was systematically excluded from multiple development opportunities over 18 months, significantly damaging career trajectory.

*Victim marginalization.* In the marginalized condition, the victim was described as a single parent with limited financial resources who relied on the position for family healthcare. In the non-marginalized condition, the victim was described as having a dual-income household with significant savings and alternative opportunities.

*Alternative availability.* In the exhausted condition, the manager had previously raised concerns through HR, filed formal complaints, and escalated to senior leadership, all without remedy. In the available condition, formal channels had not been attempted, and the organization had a reputation for taking discrimination claims seriously.

Manipulations were checked through items assessing perceived severity (3 items,  $\alpha = .91$ ), victim vulnerability (3 items,  $\alpha = .86$ ), and alternative availability (2 items,  $\alpha = .82$ ). All manipulations operated as intended (all manipulation check *t*-tests significant at  $p < .001$  with large effects,  $d_s > 1.1$ ).

#### Measures

*Classification.* Participants classified the manager's behavior using the same three-category measure as Study 1.

*Moral outrage.* Following the scenario but before classification, participants reported moral outrage using five items adapted from Batson et al. (2007): "outraged," "angry," "disgusted," "indignant," and "furious" ( $\alpha = .93$ ). Items were rated on 7-point scales (1 = *not at all*, 7 = *extremely*).

*Additional measures.* Participants also completed measures of perceived organizational support for the manager's actions, anticipated consequences, and demographic items.

## Results

### Classification Frequencies

Table 2 presents classification frequencies by condition. Overall classification proportions were similar to Study 1: 38.5% obligatory, 44.2% supererogatory, and 17.3% prohibited. The slight increase in obligatory classifications relative to Study 1 may reflect differences in scenario content or sampling variation.

**Table 2.** Classification Frequencies by Condition (Study 2).

Condition	Obligatory	Supererogatory	Prohibited	<i>n</i>
<b>Severity</b>				
Moderate	28.2%	51.3%	20.5%	78
Severe	48.7%	37.2%	14.1%	78
<b>Marginalization</b>				
Non-marginalized	32.1%	48.7%	19.2%	78
Marginalized	44.9%	39.7%	15.4%	78
<b>Alternatives</b>				
Available	30.8%	50.0%	19.2%	78
Exhausted	46.2%	38.5%	15.4%	78

### Hypothesis Tests

We tested Hypotheses 2a-c using multinomial logistic regression predicting classification from the three experimental factors. For interpretability, we focus on contrasts comparing obligatory versus supererogatory classifications (i.e., among those viewing the behavior as morally positive, what predicts viewing it as required versus optional).

*Injustice severity.* Severe injustice significantly increased the odds of obligatory (versus supererogatory) classification, OR = 2.41, 95% CI [1.28, 4.53],  $p = .006$ , supporting Hypothesis 2a.

*Victim marginalization.* Marginalized victim status significantly increased the odds of obligatory classification, OR = 1.89, 95% CI [1.01, 3.54],  $p = .046$ , supporting Hypothesis 2b.

*Alternative availability.* Exhausted alternatives significantly increased the odds of obligatory classification, OR = 1.96, 95% CI [1.04, 3.69],  $p = .038$ , supporting Hypothesis 2c.

No significant interactions emerged among the three factors (all interaction  $ps > .15$ ), suggesting additive effects.

#### *Mediation Analysis*

We tested whether moral outrage mediated the effects of severity and marginalization on obligatory classification using Hayes's (2018) PROCESS macro (Model 4). Because the outcome is categorical, we estimated mediation using logistic regression for the classification outcome, with bootstrapped confidence intervals (5,000 resamples) for indirect effects.

For severity, the total effect on obligatory classification was significant,  $b = 0.88$ ,  $SE = 0.32$ ,  $p = .006$ . The indirect effect through moral outrage was significant,  $b = 0.41$ , 95% CI [0.18, 0.71], representing 47% of the total effect. The direct effect remained significant,  $b = 0.47$ ,  $SE = 0.34$ ,  $p = .048$ , indicating partial mediation.

For marginalization, the total effect was significant,  $b = 0.64$ ,  $SE = 0.32$ ,  $p = .046$ . The indirect effect through outrage was significant,  $b = 0.25$ , 95% CI [0.06, 0.51], representing 39% of the total effect. The direct effect was not significant,  $b = 0.39$ ,  $SE = 0.34$ ,  $p = .12$ , suggesting more complete mediation for this factor.

These results provide partial support for Hypothesis 3. Moral outrage partially mediates the relationship between severity and obligatory classification and more fully mediates the marginalization effect. However, substantial direct effects (particularly for severity) indicate that outrage does not fully account for these relationships. We return to interpretation of these findings in the General Discussion.

#### *Discussion*

Study 2 demonstrates that situational characteristics systematically influence moral classification of Robin Hoodism. Observers were approximately twice as likely to classify Robin Hoodism as obligatory (versus supererogatory) when injustice was severe, victims were marginalized, and alternatives had been exhausted. These effects were additive rather than interactive, suggesting that each factor independently contributes to shifting classifications toward obligation.

Moral outrage partially mediated effects of severity and marginalization. This finding is consistent with theories suggesting that emotional responses play a role in moral judgment (Haidt, 2001), though the substantial direct effects indicate that outrage is not the sole mechanism. Other potential pathways—including perceived harm magnitude, assessments of victim desert, and judgments about manager capacity—may also connect situational factors to classification. Importantly, whether outrage-driven shifts toward obligation represent accurate moral perception or emotional bias remains an open question that our design cannot resolve.

The 17% prohibited classification rate, relatively stable across conditions, suggests that a consistent minority views Robin Hoodism as impermissible regardless of circumstances. This finding warrants further attention. Prohibited classifications may reflect principled commitment to organizational authority, concern about unintended consequences of unauthorized intervention, or deontological orientation toward rule-following. We examine the stability of various classification types under reflection in Study 3.

### **Study 3: The Effect of Moral Reflection**

#### **Method**

##### *Participants*

We recruited 125 working adults through Prolific Academic using the same inclusion criteria as previous studies. Participants received 4.00 for approximately 20 minutes of participation (effective hourly rate: 12.00). The sample was 51% female, with a mean age of 40.2 years ( $SD = 10.8$ ).

### *Design and Procedure*

Study 3 employed a within-subjects pre-post design. Participants first read a detailed Robin Hoodism scenario and provided initial classification (pre-reflection). They then completed a structured moral reflection exercise. Finally, they provided a second classification (post-reflection), along with an open-ended explanation of any change.

The scenario described a manager who used discretionary authority to remedy discrimination against a subordinate after formal channels failed. The scenario was designed to be morally complex, combining clear injustice with clear procedural violation, to maximize potential for classification revision under reflection.

### *Moral Reflection Exercise*

The reflection exercise was designed to prompt systematic consideration of relevant moral factors without biasing toward any particular classification. It consisted of four components:

1. *Perspective-taking*: Participants wrote brief responses considering the situation from the perspectives of the victim, the manager, other team members, and organizational leadership.
2. *Consideration of alternatives*: Participants listed potential alternative courses of action the manager might have taken and evaluated the likely effectiveness of each.
3. *Consequence analysis*: Participants considered potential positive and negative consequences of the manager's actions for various stakeholders.
4. *Principle articulation*: Participants completed the prompt "If I were in this manager's position, I would consider it my moral responsibility to..." and explained their reasoning.

The reflection exercise took approximately 8-10 minutes to complete. Importantly, the exercise did not advocate for any particular classification but prompted consideration of multiple perspectives and factors.

## **Results**

### *Classification Shifts*

Table 3 presents classification frequencies before and after reflection. Pre-reflection classifications closely matched Studies 1-2: 33.6% obligatory, 46.4% supererogatory, 20.0% prohibited. Post-reflection classifications shifted notably: 47.2% obligatory, 33.6% supererogatory, 19.2% prohibited.

**Table 3.** Classification Frequencies Pre- and Post-Reflection (Study 3).

Classification	Pre-Reflection	Post-Reflection	Change
Obligatory	33.6%	47.2%	+13.6 pp
Supererogatory	46.4%	33.6%	-12.8 pp
Prohibited	20.0%	19.2%	-0.8 pp

McNemar-Bowker test indicated significant overall classification change,  $\chi^2(3) = 18.74, p < .001$ . Specifically, the shift from supererogatory to obligatory was significant (McNemar test for this transition,  $p < .001$ ), while changes involving prohibited classifications were not significant.

Of the 58 participants who initially classified the behavior as supererogatory, 24 (41.4%) shifted to obligatory classification after reflection, while 32 (55.2%) maintained supererogatory classification and 2 (3.4%) shifted to prohibited. Of the 42 participants who initially classified as obligatory, 38 (90.5%) maintained this classification, 3 (7.1%) shifted to supererogatory, and 1 (2.4%) shifted to prohibited. Of the 25 participants who initially classified as prohibited, 23 (92.0%) maintained this classification post-reflection, with 2 (8.0%) shifting to supererogatory.

These results support Hypothesis 4. Reflection significantly increased obligatory classifications at the expense of supererogatory classifications, while prohibited classifications remained essentially stable.

### *Asymmetric Stability*

An important pattern emerges from these data: classification stability was markedly asymmetric. Obligatory and prohibited classifications showed high stability (90.5% and 92.0% maintained, respectively), while supererogatory classifications showed substantial instability (only 55.2% maintained). This pattern suggests that supererogatory classifications may be more tentative or context-dependent than the other categories.

This asymmetry motivates the conceptual distinction we develop in the General Discussion: between thin supererogation (apparent optionality vulnerable to revision) and thick supererogation (perceived optionality that remains stable under reflection). In these data, approximately 45% of initial supererogatory classifications displayed the instability characteristic of thin supererogation.

#### *Qualitative Analysis*

Open-ended explanations for classification changes revealed several themes among participants who shifted from supererogatory to obligatory:

*Role-based duty recognition.* Many changers cited managerial responsibilities: “Thinking about it more, as a manager you have a duty to protect your team members from unfair treatment. It’s part of the job.”

*Unique position awareness.* Others emphasized the manager’s distinctive capacity: “She was the only one who could actually do something about it. When you’re the only one who can help, I think that makes it your responsibility.”

*Formal channel failure.* Several cited exhausted alternatives: “The normal channels completely failed. At that point, I think the manager was obligated to step in somehow.”

*Reflection on inaction.* Some reconsidered what inaction would mean: “I asked myself, what if she just did nothing? That started to seem really wrong, not just ‘not heroic’ but actually wrong.”

These themes suggest that reflection surfaced considerations—role duties, unique capacity, systemic failure, costs of inaction—that shifted perceived optionality toward perceived obligation.

Among the minority who maintained supererogatory classification, common themes included respect for organizational process (“Even though the formal process failed, you can’t just bypass it whenever you think you’re right”), concern about precedent (“If managers just do whatever they think is just, that could lead to chaos”), and valuing but not requiring heroism (“What she did was admirable, but I don’t think we can demand that kind of risk-taking from everyone”).

Among those maintaining prohibited classifications, explanations emphasized rule-following (“Two wrongs don’t make a right”), authority concerns (“It’s not her place to decide how to allocate resources outside her authority”), and consequentialist worries (“This might help one person but undermine the whole system”).

#### *Discussion*

Study 3 provides evidence that structured moral reflection shifts classifications from supererogatory toward obligatory. After engaging in systematic consideration of perspectives, alternatives, and consequences, participants were significantly more likely to view Robin Hoodism as morally required rather than merely praiseworthy. Importantly, this shift was asymmetric: supererogatory classifications were uniquely unstable, while obligatory and prohibited classifications remained largely unchanged.

This asymmetry is consistent with the theoretical argument that supererogatory framing may reflect organizational context rather than settled moral judgment. Organizational discourse routinely frames justice-oriented interventions as discretionary, potentially obscuring underlying moral requirements. When observers step back from organizational framing and consider relevant moral factors systematically, many revise their initial sense that intervention was merely optional.

However, several alternative interpretations warrant consideration. First, the reflection exercise, while designed to be neutral, may have contained subtle demand characteristics favoring obligatory classification. The perspective-taking prompts, for instance, may have increased identification with the victim in ways that biased toward viewing intervention as required. Second, reflection might introduce biases rather than reduce them—perhaps unreflective judgments better track genuine moral intuitions, while reflection imports contestable moral assumptions. Third, the stability of

prohibited classifications might indicate that rule-oriented participants resisted the reflection manipulation rather than that their judgments were more considered.

We cannot definitively adjudicate among these interpretations with the present data. What we can conclude is that supererogatory classifications are more susceptible to revision under reflection than obligatory or prohibited classifications—a pattern requiring explanation regardless of which interpretation proves correct.

## General Discussion

Across three studies involving 423 participants evaluating 847 scenarios, we investigated how observers classify managerial Robin Hoodism and what factors influence these classifications. Our findings reveal a complex picture of moral judgment in organizational contexts.

### *Summary of Findings*

Study 1 established that observers most frequently classify Robin Hoodism as supererogatory (47%), followed by obligatory (32%) and prohibited (20%). This pattern supports Hypothesis 1 while revealing substantial classification diversity. Study 2 demonstrated that situational factors—injustice severity, victim marginalization, and alternative exhaustion—shift classifications toward obligation, with moral outrage partially mediating severity and marginalization effects. These findings support Hypotheses 2a-c and 3. Study 3 revealed that structured moral reflection significantly increases obligatory classifications at the expense of supererogatory judgments, supporting Hypothesis 4. Critically, this effect was asymmetric: supererogatory classifications showed unique instability under reflection, while obligatory and prohibited classifications remained largely stable.

### *Thin and Thick Supererogation: A Conceptual Distinction*

Our findings motivate a conceptual distinction between two forms of supererogatory classification. We propose the terms *thin supererogation* and *thick supererogation* to describe this distinction, while emphasizing that these terms characterize patterns in observer judgment rather than objective moral properties of actions.

*Thin supererogation* describes classifications of apparent moral optionality that are susceptible to revision under systematic reflection. In our studies, approximately 45% of initial supererogatory classifications shifted toward obligatory classification after observers engaged in moral reflection. These judgments exhibited what we term thin supererogation—a sense that the action exceeded duty that, upon further consideration, shifted toward a sense that the action was required.

*Thick supererogation* describes classifications of perceived optionality that remain stable under extended moral reflection. Approximately 55% of initial supererogatory classifications remained stable post-reflection. These judgments exhibited thick supererogation—a sense of genuine moral optionality that persisted after considering multiple perspectives, alternatives, and consequences.

We emphasize several important limitations of this distinction. First, it is descriptive rather than normative: thin supererogation describes *judgments that tend to shift* rather than *judgments that are incorrect*. We cannot establish that post-reflection judgments are more accurate than pre-reflection judgments; we can only document that they differ. Second, the distinction depends on the reflection manipulation employed. Different reflection exercises might produce different patterns of stability and change. Third, stable classifications might represent genuinely considered judgments or might represent deeply entrenched biases that reflection fails to dislodge. The distinction identifies a pattern requiring explanation; it does not itself provide that explanation.

With these caveats, the thin/thick distinction may prove useful for understanding moral judgment in organizational contexts. Organizations routinely frame potentially obligatory actions as supererogatory—through language of organizational citizenship, going above and beyond, and discretionary effort. Some of this framing may track genuine optionality (thick supererogation), while

some may obscure underlying moral requirements (thin supererogation). The distinction provides vocabulary for investigating when organizational framing reflects versus distorts moral reality.

### *Theoretical Contributions*

Our findings contribute to several theoretical conversations. First, we extend organizational justice research by examining the moral psychology of justice enactment. Prior work has documented that individuals actively create justice through their behaviors (Scott et al., 2014; Sherf et al., 2019), but has not examined how observers classify such behaviors morally. Our findings suggest that justice enactment through unauthorized means is predominantly viewed as supererogatory but that this view is unstable for many observers and responsive to both situational factors and reflective processing.

Second, we contribute to philosophical discourse on supererogation by empirically examining factors that influence supererogatory classification. While philosophers have theorized about boundary conditions between obligatory and supererogatory action (Heyd, 2019), empirical investigation of how laypeople draw these boundaries has been limited. Our findings suggest that boundaries are fuzzy, context-dependent, and susceptible to revision—consistent with qualified supererogation views that emphasize the role of circumstances in determining obligation (Archer, 2016).

Third, we contribute to the emerging literature on prosocial rule breaking by examining a specific justice-oriented variant. Morrison's (2006) foundational work established that rule-breaking can serve constructive purposes, and subsequent research has examined antecedents and consequences of such behavior. Our contribution lies in examining moral classification—how observers categorize these behaviors across obligatory, supererogatory, and prohibited categories—and identifying factors that shift classification.

Fourth, we contribute to literature on moral reasoning in organizations by documenting reflection effects on moral classification. Research on ethical decision-making has examined how reflection influences choices (Zhong, 2011), but less attention has focused on reflection effects on moral categorization. Our finding that reflection asymmetrically shifts supererogatory classifications toward obligation suggests that moral framing in organizations may not survive systematic consideration.

### *The Significance of Moral Diversity*

A striking feature of our data is the persistent presence of all three classification categories across conditions and studies. Even under conditions most favorable to obligatory classification (severe injustice, marginalized victim, exhausted alternatives, post-reflection), a substantial minority maintained supererogatory or prohibited classifications. This moral diversity deserves attention.

The 19-20% of observers who consistently classified Robin Hoodism as prohibited represent a significant minority whose views are often overlooked in discussions of workplace justice. These observers view unauthorized justice intervention as impermissible regardless of motivations—a position grounded in respect for organizational authority, concern about precedent, and commitment to procedural legitimacy. This position connects to important philosophical debates about rule of law and the limits of individual moral judgment (Rawls, 1971). Organizations developing ethics frameworks should take seriously that meaningful minorities of employees may view Robin Hoodism not merely as optional but as categorically wrong.

Similarly, the persistence of supererogatory classifications even post-reflection suggests that many observers maintain considered views that Robin Hoodism, while admirable, exceeds what can be demanded of managers. This position respects limits on moral demands, acknowledges the costs and risks of unauthorized intervention, and leaves space for moral heroism without mandating it. The pluralism revealed in our data suggests that no single classification captures unanimous moral judgment.

### *Practical Implications*

Our findings carry implications for organizational practice, though we offer these tentatively given limitations in generalizing from hypothetical scenarios to actual organizational behavior.

First, organizations may consider how they frame managerial responses to injustice. Language of discretionary behavior and “going above and beyond” may obscure situations where intervention could be viewed as morally required by many observers. Organizations might benefit from clearer articulation of managerial responsibilities regarding justice concerns within their domains—specifying when intervention is expected rather than merely permitted.

Second, formal channels for addressing injustice may warrant examination. When formal channels fail, as in our exhausted-alternatives condition, observers were significantly more likely to view informal intervention as obligatory. Organizations with credible, effective formal mechanisms may reduce pressure on managers to engage in unauthorized intervention by providing legitimate pathways to remedy.

Third, organizations might acknowledge moral diversity in ethical frameworks. Rather than assuming consensus on whether justice-motivated rule-breaking is obligatory, supererogatory, or prohibited, organizations might recognize that reasonable employees disagree and create space for diverse positions. This might include protections for those who engage in good-faith justice intervention alongside respect for those who decline to circumvent formal processes.

### *Limitations and Future Directions*

Several limitations temper interpretation of our findings. Most fundamentally, the factorial vignette method, while providing experimental control, may not capture the full complexity of actual Robin Hoodism decisions. Real managers face extended temporal unfolding, ambiguous information, personal stakes, and relationship considerations that brief scenarios cannot capture. Observer classifications of hypothetical scenarios may differ from first-person decisions or evaluations of actual cases.

Additionally, all three studies examined third-party observers, but Robin Hoodism involves first-person managers, second-party victims, perpetrators, and organizational authorities who may classify actions differently. The perspectives of those directly involved likely carry distinct features that observer studies miss. Future research should examine classification patterns across stakeholder positions.

Our samples, while comprising working adults with organizational experience, were limited to U.S. participants recruited through an online platform. Cultural variation in attitudes toward authority, procedural legitimacy, and individual versus collective responsibility likely influences Robin Hoodism classification. Cross-cultural research is needed to establish boundary conditions of our findings.

The reflection manipulation in Study 3, while designed to be balanced, may have contained subtle biases favoring obligatory classification. Future research might compare different reflection protocols to examine whether results are robust across reflection methods. Additionally, whether reflection produces more *accurate* judgments (versus merely *different* judgments) remains unclear. Research designs comparing reflection effects to expert ethical judgments or behavioral consequences might address this question.

Finally, our studies did not examine consequences of classification. How does viewing Robin Hoodism as obligatory versus supererogatory affect subsequent judgments, behaviors, and relationships? Do obligatory classifications increase willingness to engage in Robin Hoodism when in the manager role? Do they increase negative evaluations of managers who fail to intervene? Understanding downstream consequences of classification would enhance practical relevance.

## Conclusions

Managerial Robin Hoodism illuminates tensions between organizational authority and substantive justice that no formal system fully resolves. Our research documents how observers navigate these tensions through moral classification, revealing that supererogatory judgments predominate but are responsive to situational factors and reflective processing. The conceptual distinction between thin and thick supererogation—between apparent optionality that shifts under reflection and perceived optionality that remains stable—provides vocabulary for understanding how organizational contexts shape moral perception, though we acknowledge limitations of this distinction and the descriptive nature of our findings.

Organizations inevitably produce injustices that formal channels fail to remedy. How managers should respond—and how others should evaluate their responses—remains genuinely contested. Our research does not resolve this contest but illuminates its contours, documenting moral diversity in classification, identifying factors that shift judgment, and demonstrating the instability of supererogatory framing under reflection. We hope these findings contribute to more nuanced conversations about moral obligation in organizational life.

## References

- Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods, 17*(4), 351-371.
- Alvesson, M., & Willmott, H. (1992). On the idea of emancipation in management and organization studies. *Academy of Management Review, 17*(3), 432-464.
- Ambrose, M. L., & Schminke, M. (2009). The role of overall justice judgments in organizational justice research: A test of mediation. *Journal of Applied Psychology, 94*(2), 491-500.
- Archer, A. (2016). Supererogation and the professional duties of doctors. *Bioethics, 30*(8), 596-604.
- Batson, C. D., Kennedy, C. L., Nord, L. A., Stocks, E. L., Fleming, D. A., Marzette, C. M., Lishner, D. A., Hayes, R. E., Kolchinsky, L. M., & Zerger, T. (2007). Anger at unfairness: Is it moral outrage? *European Journal of Social Psychology, 37*(6), 1272-1285.
- Brink, D. O. (1994). Moral conflict and its structure. *The Philosophical Review, 103*(2), 215-247.
- Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Conlon, D. E., & Wesson, M. J. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology, 98*(2), 199-236.
- Crisp, R. (2013). Supererogation and virtue. In M. Timmons (Ed.), *Oxford studies in normative ethics* (Vol. 3, pp. 13-34). Oxford University Press.
- Galperin, B. L. (2012). Exploring the nomological network of workplace deviance: Developing and validating a measure of constructive deviance. *Journal of Applied Social Psychology, 42*(12), 2988-3025.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834.
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.
- Heyd, D. (1982). *Supererogation: Its status in ethical theory*. Cambridge University Press.
- Heyd, D. (2019). Supererogation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2019 ed.). Stanford University.
- Horgan, T., & Timmons, M. (2010). Untying a knot from the inside out: Reflections on the “paradox” of supererogation. *Social Philosophy and Policy, 27*(2), 29-63.
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kant, I. (1996). *The metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press. (Original work published 1797)
- Miller, D. (2004). Holding nations responsible. *Ethics, 114*(2), 240-268.
- Morrison, E. W. (2006). Doing the job well: An investigation of pro-social rule breaking. *Journal of Management, 32*(1), 5-28.

- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22-27.
- Rawls, J. (1971). *A theory of justice*. Harvard University Press.
- Scanlon, T. M. (1998). *What we owe to each other*. Harvard University Press.
- Scheffler, S. (1982). *The rejection of consequentialism: A philosophical investigation of the considerations underlying rival moral conceptions*. Oxford University Press.
- Scott, B. A., Garza, A. S., Conlon, D. E., & Kim, Y. J. (2014). Why do managers act fairly in the first place? A daily investigation of “hot” and “cold” motives and discretion. *Academy of Management Journal*, 57(6), 1571-1591.
- Sherf, E. N., Venkataramani, V., & Gajendran, R. S. (2019). Too busy to be fair? The effect of workload and rewards on managers’ justice rule adherence. *Academy of Management Journal*, 62(2), 469-502.
- Simmons, A. J. (1979). *Moral principles and political obligations*. Princeton University Press.
- Singer, P. (1972). Famine, affluence, and morality. *Philosophy & Public Affairs*, 1(3), 229-243.
- Tappolet, C. (2016). *Emotions, values, and agency*. Oxford University Press.
- Young, I. M. (2011). *Responsibility for justice*. Oxford University Press.
- Zhong, C. B. (2011). The ethical dangers of deliberative decision making. *Administrative Science Quarterly*, 56(1), 1-25.
- Zucker, L. G. (1987). Institutional theories of organization. *Annual Review of Sociology*, 13(1), 443-464.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.