

RESEARCH ARTICLE

How to React in the Case of Powerful Transgressors: Kill Them With Kindness or Punishment?

Petr Houdek | Štěpán Bahník | Marek Vranka | Martin Zielina

Prague University of Economics and Business, Prague, Czechia

Correspondence: Petr Houdek (petr.houdek@gmail.com)

Received: 2 June 2025 | Revised: 12 February 2026 | Accepted: 11 March 2026

ABSTRACT

When wrongdoing is committed by a powerful superior, victims face a strategic dilemma: punish (e.g., confront and report), avoid or forgive. Moral-repair and restorative-justice perspectives suggest that conciliatory responses can elicit guilt and restore cooperation, whereas punitive responses may deter misconduct but risk backlash, especially in situations of power asymmetry. Across four studies, we examined both sides of this dyadic process: how victims select response strategies and how powerful transgressors subsequently react. Study 1 ($N = 1003$) showed that observers expected more punishment for more severe supervisor transgressions, yet also anticipated that punished superiors would respond more unkindly than forgiven superiors. Two incentivized laboratory studies modelling asymmetric appropriation (Study 2a: $N = 503$; Study 2b: $N = 487$) replicated the severity–response gradient; however, the subsequent behaviour of powerful transgressors was largely unresponsive to either punishment or forgiveness. A retrospective incident survey of recent supervisor misconduct (Study 3; $N = 395$) revealed that avoidance was the most common response; higher perceived severity predicted punishment, and punishment was associated with greater supervisor retaliation and lower reconciliation. Across methods, we document an expectation–outcome gap as responses that are normatively endorsed and intuitively appealing often fail to induce moral repair when the transgressor is insulated by power.

1 | Introduction

Addressing workplace misconduct is crucial for maintaining a healthy work environment. However, existing research on counterproductive work behaviour shows that such unethical behaviours are common. Possible responses to it are highly heterogeneous, ranging from immediate organizational justice to ignoring it to supporting the transgressors, especially when high-status employees are among them (Deshpande et al. 2000; Greve et al. 2010; LaVan and Martin 2008).

Punishment is generally accepted as a fair and adequate procedure against workplace misconduct (Greve and Teh 2016). Because field experimental studies are logistically challenging, researchers have mostly relied on online and laboratory experiments. Indeed, these paradigms operationalize wrongdoing as

norm-violating self-interested taking and do not capture the full phenomenology of organizational misconduct. However, they are useful models of supervisor wrongdoing because they enable to study unfair appropriation and exploitation of resources under asymmetric control. These studies consistently show that punitive responses to unethical behaviour are viewed as the most effective, intuitive and personally satisfying—and thus the most widely expected—reactions to misconduct (Brandt et al. 2006; Duersch and Müller 2015; Fehr and Fischbacher 2004; Fehr and Gächter 2000, 2002; Galeotti 2015; Kakkar et al. 2020; Yang et al. 2025).

This retributive justice orientation is often driven by a desire for justice or retribution, aiming to reestablish fairness in relationships and penalize those who violate ethical norms (Butterfield et al. 2023; Kurzban et al. 2013). However, although

these reactions may serve an immediate purpose, they can lead to ongoing resentment, increased conflict with cycles of retaliation, and significant resources being allocated to punishing (Dreber et al. 2008; Heffner and FeldmanHall 2019). When punishment can be counter-punished, would-be sanctioners reduce punishment and cooperation deteriorates, because the expected costs of enforcing norms rise (Nikiforakis 2008). Related work demonstrates that asymmetric punishment and unequal enforcement capacities alter both sanctioning and outcomes (Nikiforakis et al. 2010, 2012). Additionally, it shows that retaliation can escalate into costly feuds that offset potential gains from cooperation (Nikiforakis et al. 2012). In the actual workplace, it is crucial to carefully consider the power status of the transgressor and the long-term consequences of punitive actions. Efforts to punish the transgressors often fail to materialize, potentially leading to work retaliation and victimization (Cortina 2008; Cortina and Magley 2003). Taken together, these findings show that punitive responses may be psychologically compelling yet strategically fragile when the victim anticipates backlash from a structurally advantaged transgressor.

As an alternative to punitive or retributive approaches, various ethical doctrines advocate a less intuitive strategy: responding to transgressions with understanding, forgiveness and kindness (Enright et al. 1998; Norlock 2022). This restorative justice orientation (Butterfield and Goodstein 2010) posits that positive, non-retaliatory responses can elicit guilt or shame in the transgressor, potentially leading to self-regulation of antisocial behaviour and fostering prosocial actions (Goodstein et al. 2016; McCullough 2001). The restorative justice approach expects the transgressor to acknowledge guilt and make amends (Berndsen and Wenzel 2021; Bottom et al. 2002; Wenzel et al. 2023). Forgiveness from the victim signals readiness to maintain the relationship, discouraging the transgressor from continuing to violate social norms and causing irreparable harm (Wallace et al. 2008).

The theory of moral repair argues that preserving organizational sense-making and trust in shared moral standards requires constructive restoration involving both the victim and the transgressor, thereby reaffirming that norm violations are condemned and ethical expectations will be upheld (Walker 2006). Yet non-retributive responses such as forgiveness may be undervalued because punishment is often treated as the appropriate signal of justice, and forgiveness can be interpreted as weakness. Moreover, in workplace hierarchies, targets' responses to supervisor wrongdoing are rarely isolated moral gestures; they are strategic and risky choices shaped by anticipated long-term consequences and the possibility of backlash (Butterfield et al. 2023; Chiang et al. 2023; Goodstein et al. 2016). At the same time, most moral-repair work has emphasized how leaders respond to subordinates' misconduct, leaving limited empirical evidence on how employees attempt moral repair after being harmed by a powerful superior and how such power holders react to punishment, avoidance or forgiveness in turn.

We report four studies on how targets respond to wrongdoing by powerful superiors (punish, ignore, and forgive) and how those superiors react (retaliate, ignore, and reconcile). Study 1 measures observers' expectations using vignettes. Studies 2a/2b use incentivized multi-round lab interactions to test actual responses

and whether punishment or forgiveness alters power-holders' behaviour. Study 3 surveys employees' recalled incidents to capture real-world responses and supervisors' reactions. Together, the studies identify what drives targets' strategy choices and when these strategies are ineffective or risky under power asymmetry.

2 | Theoretical Framework and Research Questions Development

Previous research on the theory of moral repair and the restorative justice orientation (Butterfield and Goodstein 2010; Goodstein et al. 2016; McCullough 2001; Vives-Gabriel et al. 2023) has identified factors associated with forgiveness for unethical behaviour, such as perceiving the offence as less severe, low recidivism rate of offenders, empathetic understanding or seeing oneself as similar to the transgressor (Exline et al. 2008; Fatas and Restrepo-Plaza 2022; Riek and Mania 2012). Another line of inquiry has examined whether forgiveness leads to increased cooperation or is exploited for further offences (Wallace et al. 2008). However, there is a gap in understanding how victims of transgressions in organizational settings choose to forgive, overlook or punish in strategic interactions with powerful transgressors.

This lack of understanding is particularly evident when compared to the breadth of research on destructive leadership and abusive supervision (Liu et al. 2018; Waytz et al. 2013). Similarly, a rich psychological literature demonstrates how positions of power lead to a decline in empathy, reduced trust in others and an increase in antisocial behaviour towards those without power (Foulek et al. 2018, 2020; Galinsky et al. 2006). A further risk is introduced by the selection effect, whereby it can be expected that there will be a strong connection between guilt-proneness and sensitivity to others' distress; thus, people who seek power over others will most likely have a higher threshold for guilt or shame (Ent and Baumeister 2015; Hu et al. 2024). Knowing what the appropriate strategy is for dealing with unethical behaviour by a powerful supervisor is, therefore, crucial. Because power both increases the likelihood of transgression and reduces exposure to social and formal sanctions, it also changes the strategic calculus for targets as punishment or voice may be riskier due to backlash and less effective because the superior can ignore costs, whereas forgiveness may fail to elicit guilt- or shame-driven repair when accountability is low. This power-based asymmetry is the reason we examine victims' strategy selection together with the subsequent reactions of powerful transgressors.

Employees may react to the transgressions of superiors by punishing them (such as voicing concerns, reporting the behaviour to HR or the compliance department, blowing the whistle or filing a lawsuit), ignoring them, or forgiving them (Miceli and Near 1994, 2005; Morrison 2014). In laboratory settings, researchers often examine how responders behave in Power-to-Take Game experiments, where 'employees' can punish selfish 'managers' for seizing their money (Bosman and van Winden 2002; Galeotti 2015). However, most of these experiments are one-shot and without a forgiving option, and the relational dynamics is thus not fully operationalized. In organizations, employees' decisions to act vary based on their assessment of the likelihood of successfully addressing or punishing unethical behaviour, which raises complex moral considerations (Dungan et al. 2019; Gago

2021; MacGregor and Stuebs 2014; Waytz et al. 2013). Employees are arguably more likely to exhibit risk-averse behaviour due to the uncertainty of the resolution of transgression, which often provokes negative emotions and heightens their attention to the potential drawbacks or adverse outcomes of a given action (Leana et al. 2012; Lerner and Keltner 2001; Novaro et al. 2024). Indeed, their decision-making is complicated, as they must consider the likelihood of punishment and the inherent power dynamics that create the fear of retaliation—that is, how the supervisor, who has power over them, will react to their behaviour (Cortina 2008; Cortina and Magley 2003).

2.1 | Expected Strategies Under Power Asymmetry

To explore how people expect others to respond to unethical behaviour by powerful figures, we conducted an online vignette study, allowing us to gather a broad range of responses from a large, diverse sample of working adults across various professional contexts (Study 1). This method enabled us to investigate participants' expectations about how employees would respond to unethical behaviour (punishment, forgiveness or ignoring the misconduct) and how they expect superiors to react to these responses.

RQ 1a: What are the expectations about the responses of employees to their superiors' unethical actions toward them? Do people expect that they will choose punishment, inaction or forgiveness?

Importantly, targets' strategy selection is constrained not only by what they would like to do but also by how they expect the superior to react. Research on employee voice and upward challenge reveals that powerful actors often respond defensively to being confronted, typically by discounting the message, derogating the messenger or retaliating (Detert and Burris 2007; Burris 2012). More generally, responses to confrontation can involve anger and counteraggression rather than moral self-correction, particularly when the confronted party experiences shame or threat to status (Tangney et al. 2007). These dynamics are central in asymmetric relationships because power reduces accountability and increases the feasibility of backlash, making it theoretically plausible that punishment can increase retaliation rather than induce moral repair. Thus, we ask:

RQ 1b: What are the expectations about the behavior of superiors who have acted unethically when a subordinate responds to their problematic behavior by punishing, ignoring or forgiving them?

2.2 | Victims' Strategies Under Power Asymmetry

We conducted controlled laboratory experiments (Studies 2a/b)¹ to examine the same dynamics under controlled conditions. The laboratory setting enables us to isolate specific variables of interest and manipulate power dynamics directly, allowing for an examination of how targets actually behave when confronted

with unethical behaviour from a more powerful transgressor and how transgressors subsequently respond to victims' strategies. Specifically, the experiment modelled superior–subordinate interactions using a modification of the taking version of the Dictator Game to assess the influence of punishment, indifference and forgiveness on the subsequent behaviour of powerful individuals. We then verified the results using the real incident survey (Study 3), which enhances the ecological validity of laboratory findings by capturing real incidents involving one's supervisor and documenting the victim's chosen response, the supervisor's reaction and any downstream consequences.

RQ 2a: How do employees respond to their superiors' unethical actions toward them? Do they choose punishment, inaction or forgiveness?

We sought to determine which strategy individuals employ when confronted with unethical behaviour of powerful transgressors against themselves and whether different strategies can induce positive change in some transgressors in an incentivized setting.

Responses to transgressions are influenced not only by power dynamics but also by the psychological and social attributes of both victims and transgressors. In victims, we highlight three constructs that are likely to influence the decision-making process when confronting unethical conduct in hierarchical relationships: *Political will* as an instrumentally focused construct, *empathy* as a morally oriented construct and *risk aversion* as a reflection of decision-making under uncertainty. These personal traits can influence how victims interpret and respond to the unethical behaviour of superiors, shaping whether they opt for punishment, inaction or forgiveness. Individuals with high political will are more effective in navigating group and power dynamics and selecting appropriate situational responses (Frieder et al. 2019). Political will refers to the motivation to engage in influence attempts despite the risks of interpersonal and reputational consequences (Kapoutsis et al. 2017). In the present context, punitive responses map onto forms of upward challenge (e.g., reporting, escalating and confronting) that are known to carry backlash risk from managers, particularly when the input is challenging (Detert and Burris 2007; Burris 2012). We therefore expect that individuals higher in political will be more likely to select punitive behaviour rather than avoidance when facing supervisor wrongdoing.

Empathy, on the other hand, involves the intuitive understanding and sharing of others' feelings, which can significantly influence how one perceives and reacts to moral transgressions (Bloom 2017; Erlandsson et al. 2015). An empathetic individual may be more inclined to forgive, understanding the complexities and pressures that might lead a transgressor to act unethically. Additionally, empathy contributes to emotional intelligence, which is linked to better interpersonal relations and conflict-resolution skills (Klimecki 2019).

Risk aversion, meanwhile, affects how individuals weigh the potential costs and uncertainties associated with confronting powerful transgressors. More risk-averse individuals may be hesitant to punish a superior, fearing retaliation or professional

consequences, even when the unethical behaviour is severe (Cortina and Magley 2003). This cautious approach can lead to passive strategies, such as ignoring the misconduct, as a means of self-protection. Conversely, those with lower risk aversion may be more willing to take decisive actions despite possible repercussions, particularly when fairness and justice are at stake.

RQ 2b: How are victims' political will, empathy, and risk preferences associated with a response (punishment, inaction or forgiveness) to the unethical actions of a powerful transgressor?

2.3 | Transgressor Reactions to Victims' Strategies

Whether a powerful transgressor reacts with moral repair versus defiance and retaliation should depend in part on their dispositional tendency to experience self-conscious moral emotions. We therefore focus on guilt- and shame-proneness as theoretically relevant moderators of post-confrontation (or post-forgiveness) behaviour. Guilt focuses on the wrongness of a specific action, whereas shame pertains to broader self-assessments (Tangney 1995; Tangney and Dearing 2002). Extensive prior research has consistently demonstrated a strong inverse relationship between guilt proneness and engagement in unethical actions (Cohen et al. 2012). In particular, individuals with higher levels of guilt-proneness are less likely to make unethical choices, including engaging in counterproductive workplace behaviours (Cohen et al. 2013). These emotions are painful, and they show that one cares about having wronged someone, not living up to one's morals, or having chosen flawed morals, thereby losing others' trust. To move past these feelings, one must reevaluate one's behaviour, apologize or make amends (Goodstein et al. 2016; Hareli and Eisikovits 2006).

RQ 3a: How do transgressors respond to their victims' actions toward them (punishment, inaction or forgiveness)? Do they choose retaliation, no action or reconciliation?

RQ 3b: How is the propensity to feel guilt and shame in transgressors associated with their decision-making after the victim's response?

3 | Study 1

We investigate participants' expectations about how employees respond to unethical behaviour by superiors and how those superiors might react to such responses. We explored scenarios depicting varying levels of supervisor misconduct (mild or severe). Participants decided which strategy the employee would choose (punishment, inaction or forgiveness) and anticipated subsequent reactions from superiors (conciliatory or punitive).

Theoretical frameworks suggest that responses to unethical behaviour are shaped by moral repair strategies, emphasizing retributive justice through punishment or restorative approaches involving forgiveness (Fehr and Fischbacher 2004; McCullough 2001). Punitive actions are often driven by concerns about fairness but risk escalating conflicts and fostering retaliation. Conversely,

forgiveness can promote reconciliation but may be perceived as a weakness, enabling further misconduct (Cortina 2008; Waytz et al. 2013).

We expect that more unethical behaviour of the superior will lead participants to predict that the employee will react with punishment (vs. inaction or vs. forgiveness) and with inaction (vs. forgiveness). We explore the predicted effectiveness of the employees' responses, but we do not have a specific prediction of which response is expected to lead to the kindest response from the superior.

3.1 | Methods

3.1.1 | Participants

We invited a gender-balanced sample of 1000 UK and Polish nationals,² at least 18 years old, employed, fluent in English, with the proportion of their previous Prolific submissions approved by researchers collecting data of at least 99%. Ultimately, 1003 participants completed the task (523 male, 480 female; $M_{\text{age}} = 34.4$, $SD_{\text{age}} = 10.2$). See [Supporting Information](#) section for detailed information.

3.1.2 | Procedure and Materials

Each participant responded to 17 vignettes in a random order: 16 vignettes depicted instances of unethical behaviour exhibited by a supervisor towards a subordinate. In addition, we included one laboratory benchmark vignette that described the same incentive-compatible interaction used in Studies 2a/2b (a modified taking dictator-game paradigm).

The workplace scenario vignettes were designed based on typologies of deviant workplace behaviours and counterproductive work behaviour (Robinson and Bennett 1995), as well as violations of moral foundations (Graham et al. 2011). We pretested them on several independent samples of participants to verify that the vignettes were perceived as realistic and established morally questionable behaviour in the workplace (see [Supporting Information](#) section). Specifically, scenarios addressed fairness (e.g., fairness in awards assignments or resource allocation), betrayal, social exclusion, oppression, degradation, subversion, idea theft, impossible ultimatums, harassment, scapegoating or undermining feedback.

We manipulated the intensity of unethical behaviour in the vignettes, with each vignette having two levels of unethical behaviour towards a subordinate. To minimize the influence of a possible dominant choice, the vignettes randomly presented two out of three possible responses of a subordinate (ignore, forgive or punish; see further elaboration in Study 2A), and the participants were asked to select the response they believed the subordinate would choose. Next, participants were asked to imagine that the subordinate had chosen one of the two presented responses (randomly determined) and were asked to predict the supervisor's reaction to this subordinate's response. For the 16 vignettes, participants answered on a 6-point Likert scale ranging

A team leader publicly commits to backing a team member's innovative project proposal during a high-level meeting. However, behind closed doors, the leader **does not speak enthusiastically about the project, which makes the project's future unclear leading to its rejection**. The team member, feeling betrayed, considers the following responses:

Ignore the leader and continue working on other projects while maintaining professionalism, despite the setback.

Punish the team leader by disclosing serious errors in the team leader-driven projects.

The team member knows that, after some time, the leader will be in a position to recommend the team member for a prestigious award. What do you expect the team member will do?

Ignore

Punish

Imagine that the team member decides to **ignore** the leader.

Subsequently, the leader is in a position to recommend the team member for a prestigious award. Rate on the scale of what behavior you expect from the leader.

The leader's behavior towards the subordinate will be:

extremely kind	very kind	somewhat kind	somewhat unkind	very unkind	extremely unkind
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE 1 | Example vignette with instructions for the less-severe level of unethical behaviour (the more-severe version included bolded text: 'disparages the project and slanders its author to senior management, leading to its rejection') and the ignore–punish condition (out of ignore, forgive or punish). Text below the line was displayed only after selecting one of the two response options.

from 'extremely kind' (1) to 'extremely unkind' (6). An example vignette is shown in Figure 1.

To obtain expectations of behaviour in Studies 2a/b, one vignette described unethical behaviour exhibited by a player in the role of dictator towards a subordinate player in a laboratory experiment. The procedure for participants was identical; however, for this laboratory vignette, they were asked to indicate how much they believed the dictator would take in the second round. The vignette is shown in Figure 2.

During the study, two questions were used to check whether participants remembered the content of the question and were paying attention. Participants received an extra reward for answering both checks correctly.

3.2 | Results

Other preregistered analyses, along with a full description of the empirical strategy for all studies, are presented in the [Supporting Information Results](#) section.

3.2.1 | Workplace Scenario Vignettes

Participants were more likely to expect that the subordinate would forgive the superordinate when the infraction was less severe when the alternative was to ignore, $t(15.4) = -4.10$, $p < 0.001$, $b = -0.09$, 95% CI $[-0.13, -0.05]$, as well as when the alternative was to punish, $t(14.8) = -8.74$, $p < 0.001$, $b = -0.19$, 95% CI $[-0.23, -0.15]$, and more likely to think that they would punish the superordinate rather than ignore them if the infraction was more severe, $t(14.7) = 8.00$, $p < 0.001$, $b = 0.16$, 95% CI $[0.12, 0.19]$ (see Figure 3).

Participants expected the supervisor to behave more unkindly when their infraction was more severe (see Table 1 for the full results of the analysis). Moreover, they expected the supervisor's behaviour to be more unkind when the subordinate responded with punishment and more kind when the subordinate responded with forgiveness. The significant interaction for the forgive–punish and ignore–punish conditions shows that the effect of the response is slightly stronger when the supervisor's infraction is less severe, but it is in the same direction and significant for both levels of severity (see [Supporting Information Results](#) section for a graph of the interaction).

This is the beginning of this study. Below is the first vignette, with more to come.

Participants in a study played an experimental game. One was Player A, and the other was Player B. Players A and B received £20 each. Player A then had the option to take between £0 and £10 from Player B. Player A took ~~£2~~ **£10** from Player B. Player B considered the following offered options for response:

Punish Player A. From a symbolic £0 to £10. For every £1 Player B spends on punishment, Player A loses £1 as well.

Forgive Player A and send him or her additional money from £0 to £10. For every £1 Player B sends from their own money, Player A receives £1.

Player B knows that both Players will again receive £20 in the next round. Again, Player A will be able to take money from Player B, this time from £0 to £20. What do you expect Player B will do?

- Punish**
- Forgive**

Imagine that Player B decided to **punish** Player A.

In the next round, Players A and B again received £20. Player A once again had the option to take money from Player B, this time from £0 to £20. What do you think Player A will do? How much do you expect him to take from Player B?

- £0
- £2
- £4
- £6
- £8
- £10
- £12
- £14
- £16
- £18
- £20

FIGURE 2 | Vignette described unethical behaviour exhibited by a player in the role of dictator towards a player in the role of receiver in a laboratory experiment and the punish–forgive condition (out of ignore, forgive and punish). Text under the line was displayed only after selecting one of two options.

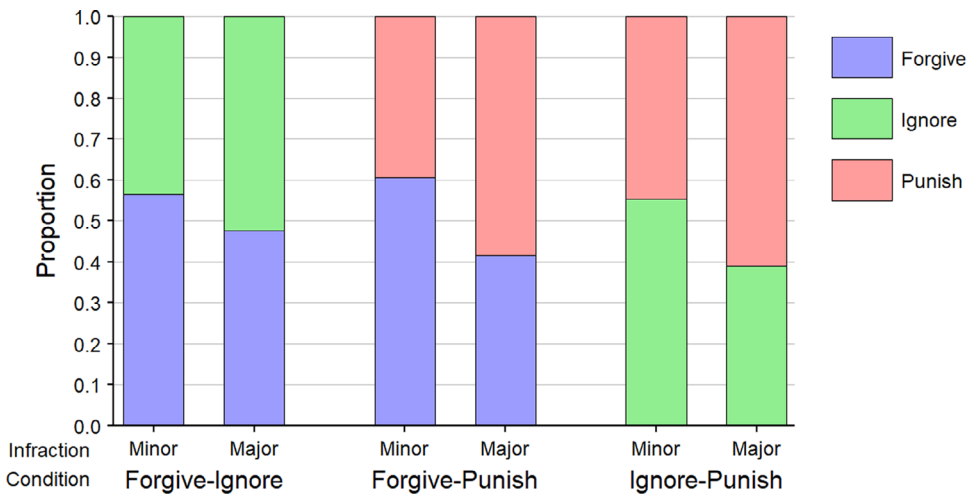


FIGURE 3 | The proportion of participants expecting the three possible responses by condition and infraction severity.

3.2.2 | Laboratory Vignette

As preregistered, we report the laboratory benchmark vignette separately because it is designed to bridge Study 1 expectations with the incentive-compatible interaction used in Studies 2a/2b.

A linear regression with the response as a dependent variable and withdrawal size as a predictor showed that for the laboratory vignette, participants believed that the receiver would be more likely to punish the dictator when the dictator took more money both when the alternative was to forgive the dictator,

TABLE 1 | The results of linear mixed-effect models of expected supervisor's behaviour (unkindness) by condition.

Predictors	Forgive–ignore		Forgive–punish		Ignore–punish	
	Estimates	<i>p</i>	Estimates	<i>p</i>	Estimates	<i>p</i>
(Intercept)	3.43 (3.28–3.58)	<0.001	3.87 (3.77–3.96)	<0.001	4.11 (4.00–4.21)	<0.001
Severity	0.26 (0.14–0.37)	<0.001	0.21 (0.14–0.28)	<0.001	0.24 (0.14–0.33)	<0.001
Response (punish or ignore)	0.47 (0.40–0.55)	<0.001	1.15 (0.97–1.32)	<0.001	0.74 (0.51–0.98)	<0.001
Severity × Response	0.01 (−0.11 to 0.13)	0.859	−0.20 (−0.31 to −0.08)	0.001	−0.14 (−0.25 to −0.03)	0.015
Marginal <i>R</i> ² /Conditional <i>R</i> ²	0.058/0.349		0.205/0.449		0.115/0.320	

Note: The response is coded as 0.5 for punishment (or ignoring in the forgive–ignore condition) and −0.5 for forgiveness (or ignoring in the ignore–punish condition). The less forgiving (more punishing) response is thus coded as 0.5. Severity is coded as −0.5 (less severe infraction) and 0.5 (more severe infraction). The numbers in parentheses show 95% confidence intervals of the estimates. The dependent variable is a rating of unkindness on a 6-point Likert scale. Boldface indicates statistically significant effects at the 5% level ($p < 0.05$).

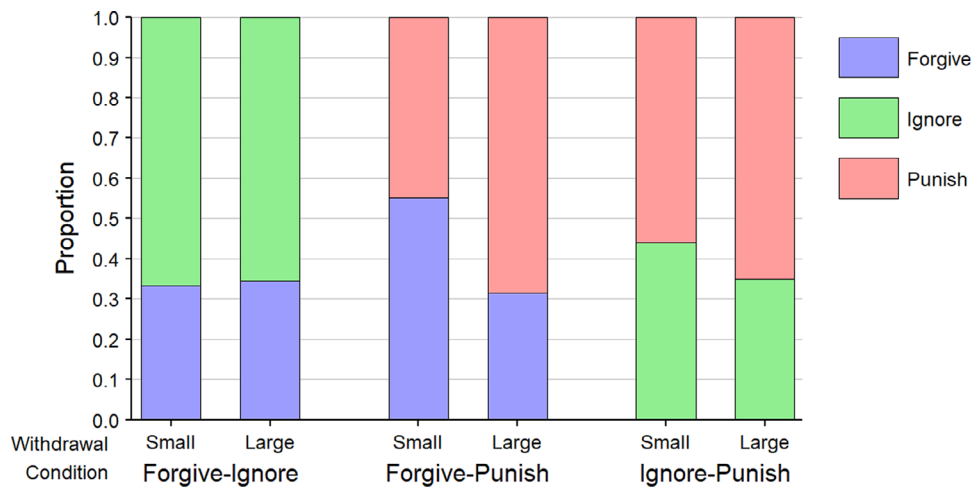


FIGURE 4 | The proportion of participants expecting the three possible responses by condition and withdrawal size in the laboratory vignette.

$t(338) = -4.49, p < 0.001, b = -0.24, 95\% \text{ CI} [-0.34, -0.13]$, and when the alternative was to ignore him, $t(334) = 1.74, p = 0.084, b = 0.09, 95\% \text{ CI} [-0.01, 0.20]$, even though only the former effect was significant. There was no significant difference in the expected response based on the withdrawal size when the alternatives were to forgive or ignore the dictator, $t(325) = 0.23, p = 0.818, b = 0.01, 95\% \text{ CI} [-0.09, 0.12]$ (see Figure 4).

Participants expected the dictator to take more money in the second round if they had taken more in the first round (see Table 2). Furthermore, they expected that dictators would take more if the recipient ignored them rather than forgave them and if they ignored them rather than punished them. The significant interaction in the forgive–ignore condition shows that the effect of the response is present only when the dictator took a larger amount of money, $t(163) = 4.74, p < 0.001, b = 4.64, 95\% \text{ CI} [2.71, 6.57]$, whereas it is not significant and close to zero when the dictator took a smaller amount of money, $t(160) = 0.02, p = 0.986, b = 0.01, 95\% \text{ CI} [-1.67, 1.70]$.

3.3 | Discussion

Findings of Study 1 reveal that participants expect employees to respond somewhat differently based on the severity of the

unethical behaviour. Participants were more likely to anticipate employees to choose punishment and less likely to choose forgiveness when superiors engaged in more severe unethical actions. This aligns with theories of justice and retribution, where individuals intuitively and emotionally seek to restore fairness and justice through corrective actions (Bosman et al. 2005; Rupp and Bell 2010).

The results further indicate the prevailing beliefs that the nature of the subordinate's response can amplify or mitigate the supervisor's behaviour, particularly in less severe infractions (Amore et al. 2023; Jadaszewski et al. 2024; Watkins et al. 2015). Specifically, superiors were expected to behave more unkindly when the subordinate chose punishment and more kindly when the subordinate opted for forgiveness. The expectation of kindness in response to forgiveness is supported by theories of emotional regulation and the supervisory alliance, which emphasize the importance of repairing relational ruptures through positive interactions (Watkins et al. 2015).

Interestingly, for the laboratory vignette, participants expected that both punishment and forgiveness would elicit a more favourable response from the dictator. Yet, as shown by the direct comparison of the two responses, they believed that forgiveness would be more effective. There are several important differences

TABLE 2 | The results of linear regression models of expected withdrawals by the dictator by condition in the laboratory vignette.

<i>Predictors</i>	Forgive-Ignore		Forgive-Punish		Ignore-Punish	
	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>
(Intercept)	9.79 (9.16 – 10.43)	<0.001	9.42 (8.67 – 10.17)	<0.001	10.08 (9.36 – 10.79)	<0.001
Severity	7.19 (5.92 – 8.47)	<0.001	5.99 (4.49 – 7.50)	<0.001	6.56 (5.14 – 7.98)	<0.001
Response (punish or ignore)	2.33 (1.05 – 3.60)	<0.001	1.19 (-0.31 – 2.70)	0.120	-1.90 (-3.32 – -0.48)	0.009
Severity x Response	4.62 (2.07 – 7.18)	<0.001	-0.26 (-3.27 – 2.75)	0.863	-0.06 (-2.91 – 2.78)	0.964
R ² / R ² adjusted	0.315 / 0.309		0.157 / 0.149		0.219 / 0.212	

Note: The response is coded as 0.5 for punishment (or ignoring in the forgive–ignore condition) and -0.5 for forgiveness (or ignoring in the ignore–punish condition). The less forgiving (more punishing) response is thus coded as 0.5. Severity is coded as -0.5 (small withdrawal) and 0.5 (large withdrawal). The numbers in parentheses show 95% confidence intervals of the estimates. The dependent variable represents the expected amount of money withdrawn in GBP. Boldface indicates statistically significant effects at the 5% level ($p < 0.05$).

between the laboratory vignettes and the remaining vignettes that could explain the results. In the workplace vignettes, an established unequal relationship exists between the two parties involved, and it is expected to continue in the future. On the other hand, in the laboratory vignette, the relationship is not explicitly described, but it can be assumed that the differences in roles are not based on merit and are purely accidental; moreover, the relationship would not last beyond the artificial situation.

These findings reveal a paradox; although participants saw punishment as likely, they also expected it to backfire via superior retaliation, reflecting awareness of the subordinate–superior power imbalance. This helps explain why employees rarely move quickly to penalize unethical supervisors and instead choose cautious, risk-averse responses, given that outcomes of reporting misconduct are uncertain (Kiewitz et al. 2016; Leana et al. 2012; Lerner and Keltner 2001; Novaro et al. 2024).

One way to reconcile the paradox is to distinguish moral endorsement from strategic feasibility. In settings where retaliation is possible, punishment deters only when it is not met by counter-punishment; when counter-punishment is feasible, sanctioning declines and enforcement can devolve into costly feuds (Nikiforakis 2008; Nikiforakis et al. 2012). Thus, subordinates may endorse punishment as appropriate yet expect backlash rather than repair from powerful transgressors. Stated support for punishment in vignettes may also reflect reputational or identity incentives, as punishing can signal trustworthiness and confer reputational benefits even when costly (Jordan et al. 2016). More generally, vignette judgments need not translate into behaviour under real incentives and risks, consistent with the intention–action gap in whistleblowing (Oelrich 2021).

Although the results from Study 1 were consistent across both British and Polish participants and a range of scenarios described in the vignettes, the judgments and reported attitudes were elicited using hypothetical vignettes (Doliński 2018; Vranka and Houdek 2024). Controlled laboratory experiments reported as

Studies 2a/b address this limitation. They simulate hierarchical interactions and introduce financial incentives, allowing us to explore whether participants' reactions would align with their stated expectations when facing material consequences.

4 | Study 2

In addition to exploring how subordinates respond to their superiors' unethical actions towards them in a controlled laboratory setting, Study 2 also aims to identify personality characteristics influencing victim–transgressor interactions. Using a modified Dictator Game paradigm, it models a stylized form of supervisor (Dictator) wrongdoing (unfair appropriation of a subordinate's resources under asymmetric control) and tests how victims' (Receiver) responses (punishment, forgiveness or neutrality) shape subsequent behaviour by the powerful transgressors. By simulating power dynamics and incentivizing decisions, the experiment examines whether responses promote moral repair and reinforce accountability.

We hypothesize that, as in Study 1, recipients' responses will depend on the severity of the transgression, with more severe actions prompting higher punishment and lower forgiveness (Fehr and Fischbacher 2004; McCullough 2001). Punishment may escalate unethical behaviour by transgressors, whereas forgiveness could reduce it (Dreber et al. 2008; Fudenberg et al. 2012). Notably, evidence from Study 1 (the laboratory vignette) also suggests that participants perceive forgiveness as a more effective strategy than punishment in this task.

Emotional dispositions or transgressors, such as a tendency to feel guilt and shame, are expected to moderate these effects (Cohen et al. 2012; Tangney 1995). Victims' personality traits, including political will, empathy and risk-aversion, are anticipated to shape their likelihood of choosing forgiveness or punishment (Frieder et al. 2019; Klimecki 2019; Leana et al. 2012; Lerner and Keltner 2001). In addition to these traits, we also measured current affective states to evaluate the effects of decisions in the task on both victims and transgressors.

Studies 2a and 2b examine the same power-asymmetric interaction while varying the response menu available to the subordinate. In Study 2a, receivers were randomly offered only two of the three possible responses (punish, ignore and forgive) as in Study 1. This restricted menu enables clean pairwise comparisons, even if one option is rarely chosen in a three-way menu, and reduces susceptibility to contextual effects such as the disproportionate selection of a middle option. In Study 2b, participants in the role of Receivers could choose among all three responses, increasing external validity. Together, Study 2a provides stronger identification of pairwise preferences and choice-set effects, and Study 2b tests whether these patterns generalize when all options are simultaneously available.

5 | Study 2a

5.1 | Methods

5.1.1 | Sample

We recruited 512 participants for laboratory sessions that included the present study from a Czech business school participants' pool. Data from nine participants were incomplete due to technical errors, and their data were used where available. Out of the 503 participants for whom demographic information was available, 47.9% were women, 88.7% were students (53.6% of whom were in the field of economics or management and 19.3% in humanities or social sciences), and their median age was 22 (IQR = 4).

5.1.2 | Procedure

The study was administered in a session after an unrelated study. The experimental sessions were conducted simultaneously with groups of 4 to 20 participants, each seated separately.

To model the unequal relationship between a powerful superior and an employee, we used a taking version of the Dictator Game (List 2007; Zhang and Ortmann 2014), which was played in two rounds with communication. Participants were randomized in the role of the dictator or the receiver. The dictator and receiver pair and their roles stayed the same for both rounds. Participants had complete information about the possible choices of both players in all rounds (i.e., randomly paired two options available). The course of the whole experiment was explained to them in advance, and their understanding of it was checked by a comprehension survey. Full message wording is provided in the [Supporting Information](#) section.

5.1.3 | The First Round

Participants in both roles received 20 CZK (approx. 1 EUR). The dictator had the option to take 0–10 CZK from the recipient.

Receivers were randomly assigned to one of three conditions, which differed in the two options from the three response strategies available to them: doing nothing, punishing or forgiving. Each choice was accompanied by an option of two message variations conveying the same strategic message to the dictator (punish, forgive or do nothing) but differing in length.

We included two semantically equivalent message versions for each response to reduce the likelihood that results hinge on idiosyncratic wording or tone and increase external validity by allowing participants to communicate the same behavioural choice in either a brief, administrative form (i.e., a label) or a more relationally elaborated form. In the analyses, we collapsed across message versions and treated them as the same response strategy.

Using the strategy method, receivers indicated their response for each possible first-round withdrawal (0–10 CZK in steps of 2), whereas dictators reported expectations only for the withdrawal they actually chose.³ Both players reported their moods. Then, the decision was implemented, and the dictator received the message, along with punishment or additional money based on the receiver's decision. The task continued with the second round.

5.1.4 | The Second Round

We operationalized the power of the superior over the employee by ensuring that the recipients had no opportunity to respond to the dictator's final decision. Thus, in the second round, both participants got 20 CZK. The dictator had the option to take 0–20 CZK from the recipient. The dictator reported their mood. The recipient had no decision to make and just filled in the mood scale.

Then, the dictator's decision was implemented, and the participants continued the study with a set of questionnaires and a payout.

5.1.5 | Measures

Political will was measured with the eight-item Political Will Scale (Kapoutsis et al. 2017; 7-point Likert scale), capturing self-serving and benevolent political will, which were averaged into a single score due to high intercorrelation. *Guilt-proneness* and *shame-proneness* were assessed using selected scenarios from the Test of Self-Conscious Affect-3 (Tangney and Dearing 2002), with seven paired responses rated on a 5-point likelihood scale and averaged separately. Affect was measured using selected items from the Positive and Negative Affect Schedule (Crawford and Henry 2004), with items rated on a 5-point scale and aggregated into mean *positive* (proud, excited and determined) and *negative affect* (angry, upset, guilty, hostile, ashamed and afraid) scores. *Empathy* was measured with 16 items from the Toronto Empathy Questionnaire (Spreng et al. 2009), rated on a 5-point frequency scale and averaged. *Risk aversion* was assessed using a simplified Holt–Laury risk-elicitation task (Holt and Laury 2002; Teubner et al. 2015), with risk aversion operationalized as the number of safe choices across five decisions.

5.2 | Results

5.2.1 | Dictator's Behaviour

5.2.1.1 | First Round. The dictators withdrew on average 2.94 CZK (SD = 3.36) in the first round from the recipients; 41.7% withdrew nothing, and 11.1% withdrew the full 10 CZK. The

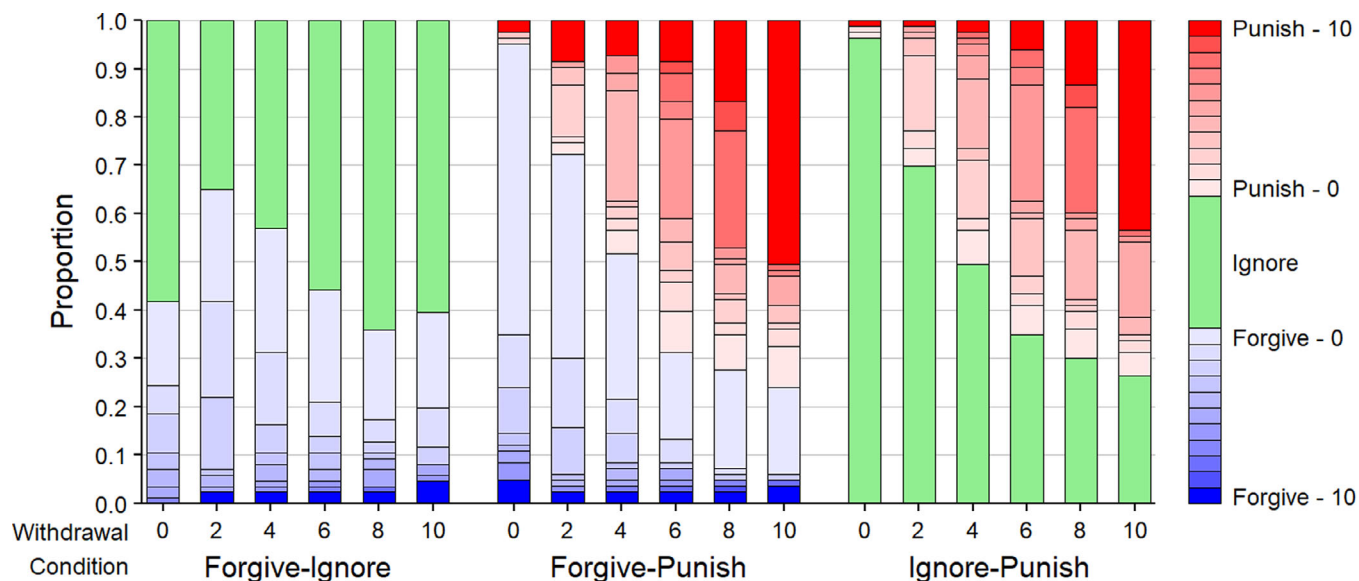


FIGURE 5 | The proportion of participants choosing the three possible responses by condition and withdrawal size.

withdrawal did not significantly differ between the three pairs of recipient's possible choices, $F(2, 249) = 1.374$, $p = 0.255$.

5.2.1.2 | Second Round. The dictators withdrew, on average 7.10 CZK ($SD = 8.15$) in the second round from the recipients. The distribution of withdrawals was bimodal: 43.3% of dictators withdrew nothing and 23.4% withdrew the full 20 CZK. The withdrawal in the second round was strongly correlated with the withdrawal in the first round, $r(250) = 0.54$, 95% CI [0.44, 0.62], $p < 0.001$.

A linear regression analysis with the withdrawal in the second round as the dependent variable, recipient's response as the predictor of interest and condition and withdrawal in the first round as covariates showed that the recipient's response did not significantly affect the withdrawal in the second round. In comparison to ignoring the dictator, neither punishment, $t(80) = 0.92$, $p = 0.362$, $b = 2.199$, 95% CI [-2.576, 6.973], nor its size (with other responses coded as punishment of size 0), $t(79) = 1.30$, $p = 0.198$, $b = 0.457$, 95% CI [-0.244, 1.158], nor forgiveness, $t(83) = -0.02$, $p = 0.983$, $b = -0.029$, 95% CI [-2.738, 2.680], nor the size of a gift, $t(83) = -0.79$, $p = 0.434$, $b = -0.258$, 95% CI [-0.911, 0.395], had a significant effect on withdrawal in the second round. Contrary to our predictions, the effects of responses were not moderated by the shame-proneness and guilt-proneness of the dictators, $ps > 0.29$ for all interaction effects. The results of the analyses were virtually unchanged when we analysed whether the dictator withdrew any positive amount using a logistic regression (see [Supporting Information Results](#) section). The only minor exceptions were that the effect of punishment size was significant, $p = 0.012$, and the interaction between punishment and shame-proneness was not, $p = 0.064$.

5.2.2 | Recipient's Behaviour

5.2.2.1 | Response. The proportion of participants choosing each response is shown in Figure 5. A mixed-effects linear

regression model with the response as the dependent variable, condition as a covariate and withdrawal as the predictor of interest, along with random intercepts and slopes for participants, was conducted to assess the association between withdrawals and responses. The withdrawal of 0 was not used in the analysis because it is not clear what is punished or forgiven in that case. The recipients were more likely to punish the dictator the more the dictator would take, $t(165.0) = 10.73$, $p < 0.001$, $b = 0.057$, 95% CI [0.046, 0.067]. The size of the punishment also increased with higher potential withdrawal by the dictator, $t(165.0) = 13.89$, $p < 0.001$, $b = 0.595$, 95% CI [0.511, 0.678]. The recipients were less likely to forgive the dictator the more the dictator would take, $t(168.0) = -8.53$, $p < 0.001$, $b = -0.048$, 95% CI [-0.059, -0.037]. The size of the gift was somewhat smaller for higher potential withdrawals, but the effect was not significant, $t(168.0) = -1.76$, $p = 0.080$, $b = -0.027$, 95% CI [-0.057, 0.003]. Contrary to our prediction, recipients were not significantly more likely to forgive that the dictator would take money if the alternative was to punish than if the alternative is to do nothing, $t(167.0) = -1.04$, $p = 0.299$, $b = -0.061$, 95% CI [-0.175, 0.054].

5.2.2.2 | Correlates of Responses. Recipients higher in political will were not more likely to forgive the dictator, nor were they less likely to punish the dictator, when the alternative was to do nothing. When both conditions where punishment was possible were included in the analysis, the average punishment size was higher for recipients higher in political will, $t(163.0) = 2.02$, $p = 0.045$, $b = 0.308$, 95% CI [0.010, 0.607].

Recipients with higher empathy were more likely to forgive the dictator when the alternative was to do nothing, $t(84.0) = 2.84$, $p = 0.006$, $b = 0.211$, 95% CI [0.065, 0.356]. Empathy was not significantly associated with the likelihood or severity of punishment in any condition.

Risk aversion showed no significant associations with forgiveness or punishment decisions. Finally, post-decision affect was not significantly related to the frequency of punishment or forgiveness

responses. See [Supporting Information Results](#) section for the full analysis of correlates of responses.

6 | Study 2b

6.1 | Methods

6.1.1 | Sample

In total, 487 participants took part in the study. Due to technical and administrative errors, data from six participants were incomplete. See [Supporting Information](#) section for detailed information. Out of the 485 participants for whom the demographic information was available, 46.2% were women and 53.2% were men, 94.0% were students (67.8% of them in the field of economics or management and 11.6% in technology or computer science fields), and their median age was 22 (IQR = 3).

6.1.2 | Procedure and Measures

The study was administered in a session after an unrelated study. The experimental sessions were conducted simultaneously with groups of 9 to 21 participants, each seated separately. The study was conducted in the same manner as Study 2a. However, participants in the role of player B had a choice of all three response strategies, rather than a selection of two. We used the same measures as in Study 2a, with the exception of a measure of risk aversion, where the choices used slightly different percentages and rewards.

6.2 | Results

6.2.1 | Dictator's Behaviour

6.2.1.1 | First Round. The dictators withdrew on average 3.39 CZK (SD = 3.78) in the first round from the recipients; 41.0% withdrew nothing and 18.4% withdrew the full 10 CZK.

6.2.1.2 | Second Round. The dictators withdrew, on average 8.14 CZK (SD = 8.59) in the second round from the recipients. The distribution of withdrawals was bimodal: 38.9% of dictators withdrew nothing and 29.9% withdrew the full 20 CZK. The withdrawal in the second round was strongly correlated with the withdrawal in the first round, $r(242) = 0.47$, 95% CI [0.37, 0.57], $p < 0.001$.

A linear regression analysis with the withdrawal in the second round as the dependent variable, recipient's response as the predictor of interest and withdrawal in the first round as covariates showed that the recipient's response did not significantly affect the withdrawal in the second round. In comparison to ignoring the dictator, neither punishment, $t(240) = 1.19$, $p = 0.236$, $b = 1.599$, 95% CI [-1.050, 4.249], nor forgiveness, $t(240) = -1.34$, $p = 0.181$, $b = -1.564$, 95% CI [-3.860, 0.732], had a significant effect on withdrawal in the second round. A similar model using punishment and gift size showed no significant effect of gift size, $t(240) = -0.16$, $p = 0.873$, $b = -0.048$, 95% CI [-0.638, 0.542]. Although higher punishment was associated with larger

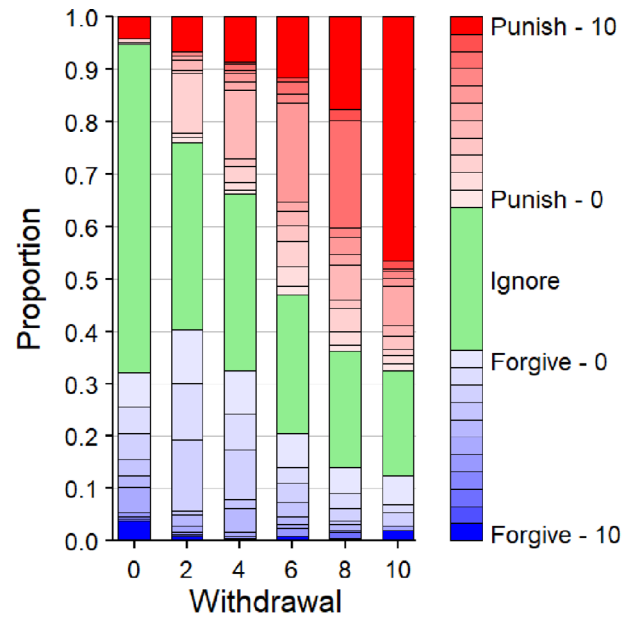


FIGURE 6 | The proportion of participants choosing the three possible responses by withdrawal size.

withdrawal in the second round, $t(240) = 1.90$, $p = 0.059$, $b = 0.307$, 95% CI [-0.012, 0.625], the effect was not significant. Contrary to our predictions, the effects of responses were not moderated by guilt-proneness of the dictators, $ps > 0.55$, for both interaction effects. However, punishment led to higher withdrawals for dictators with higher shame-proneness, $t(236) = 2.23$, $p = 0.027$, $b = 3.823$, 95% CI [0.440, 7.206]. To examine the interaction, we conducted separate regressions predicting withdrawal in the second round after punishment and inaction, including withdrawal in the first round as a covariate. Although shame-proneness was positively associated with withdrawal in the second round after punishment, $t(66) = 1.11$, $p = 0.272$, $b = 1.463$, 95% CI [-1.172, 4.099], the association was negative for participants who were ignored, $t(100) = -1.72$, $p = 0.088$, $b = -2.025$, 95% CI [-4.361, 0.310]; neither of the effects was however significant. The interaction between forgiveness and shame-proneness was not significant, $t(236) = 1.08$, $p = 0.282$, $b = 1.773$, 95% CI [-1.467, 5.012]. In an exploratory analysis, we compared the difference between withdrawal in the second round following forgiveness and punishment, controlling for the withdrawal in the first round. Participants withdrew more following punishment than following forgiveness, $t(137) = 2.34$, $p = 0.021$, $b = 3.242$, 95% CI [0.503, 5.981].

6.2.2 | Recipient's Behaviour

6.2.2.1 | Response. The proportion of participants choosing each response is shown in Figure 6. A mixed-effect linear regression with a response as the dependent variable, withdrawal as the predictor of interest and random intercepts and slopes for participants was conducted to assess the association between withdrawals and responses. The withdrawal of 0 was not used in the analysis because it is not clear what is punished or forgiven in that case. The recipients were more likely to punish the dictator the more the dictator would take, $t(242.0) = 12.92$, $p < 0.001$,

$b = 0.059$, 95% CI [0.050, 0.068]. The size of the punishment also increased with higher potential withdrawal by the dictator, $t(242.0) = 15.94$, $p < 0.001$, $b = 0.579$, 95% CI [0.508, 0.650]. The recipients were less likely to forgive the dictator the more the dictator would take, $t(242.0) = -9.17$, $p < 0.001$, $b = -0.037$, 95% CI [-0.045, -0.029]. The size of the gift was also smaller for higher potential withdrawals, $t(242.0) = -3.85$, $p < 0.001$, $b = -0.052$, 95% CI [-0.078, -0.025].

6.2.2.2 | Correlates of Responses. Political will was not significantly related to the size of punishment. Empathy was associated with less severe punishment, with more empathetic participants assigning lower punishment amounts, $t(241.0) = -2.63$, $p = 0.009$, $b = -0.886$, 95% CI [-1.546, -0.226]. Risk aversion was not significantly associated with the probability of punishing the dictator. Punishment frequency was not related to more negative affect, whereas choosing forgiveness more often was associated with more positive affect after decisions were made, $t(242) = 3.25$, $p = 0.001$, $b = 0.139$, 95% CI [0.055, 0.223]. See [Supporting Information Results](#) section for the full analysis of correlates of responses.

6.3 | Discussion of Study 2a/b

The Study 2a/b's findings suggest that punitive responses, whereas the norm in combating antisocial behaviour, are ineffective in curbing unethical behaviour if the transgressor has continued power over the victim–punisher. This ineffectiveness may stem from the power dynamics that reinforce the transgressor's ability to retaliate or maintain dominance, rendering punitive measures insufficient to alter their behaviour (Zhang and Wei 2024). Moreover, research indicates that punitive actions can exacerbate negative behaviours by fostering resentment or further unethical behaviour by the transgressor (Pleasant and Barclay 2018).

Similarly, forgiving responses, although potentially beneficial in restarting cooperation and reducing conflict, did not have an impact on powerful individuals. This may be because the anticipated mechanisms—such as eliciting guilt or shame in the transgressor—did not moderate the responses' effects due to the absence of these emotions in the transgressor's mind, rendering them ineffective in shaping behaviour. Moreover, Study 2b yielded tentative evidence that punishment may increase subsequent withdrawals among transgressors higher in shame-proneness, consistent with a defensive or retaliatory response to being confronted.

Feelings of guilt can sometimes motivate transgressors to engage in reparative actions, such as offering apologies or making amends, which are essential for moral repair (Woodyatt et al. 2022). However, defensiveness and disengagement are possible and common reactions to transgressions (Bandura 1999; Moore 2015; Kazarovytska and Imhoff 2022). Transgressors respond with disengagement or anger rather than guilt or shame, thereby resisting attempts at moral repair and perpetuating unethical behaviour.

The expectations of participants in Study 1 regarding the effectiveness of punishment and forgiveness, therefore, did not align with

the actual behaviour of participants in the laboratory experiment. The behaviour of participants with power was not malleable in a predictable manner in response to victims' actions and was consistently unethical.

On the victim's side, in Study 2a, we showed that participants with higher empathy were more likely to forgive the transgressor, but only when the alternative was to ignore them. In Study 2b, empathy was associated with less severe punishment, and choosing forgiveness was associated with more positive post-decision affect. Regardless of the costly empathetic response, the supervisors continued to behave unethically (but less than when they were punished). These findings align with research indicating that empathy or moral concern for transgressors can reduce their accountability by fostering perceptions of remorse and prompting leniency towards wrongdoers, thereby letting them 'off the hook' (Khan and Howe 2021).

In summary, our results cast doubt on the effectiveness of moral repair or retributive justice approaches in situations where powerful individuals are free to act without pressure from stakeholders or other parties (Goodstein et al. 2016; Nockur et al. 2021; Walker 2006).

The null results may be due to the nature of the laboratory situation, where the stakes are low and do not elicit a strong emotional response, or the participants may misinterpret the experimental task (Frollová et al. 2021). On the other hand, other behaviours in the experiment are consistent with our assumptions and the findings of similar studies (Engel 2011; Fehr and Gächter 2000, 2002). The victims were more likely to punish the transgressors the more they misbehaved (i.e., they withdrew more money), and the size of the punishment also increased with the higher withdrawal by the transgressor. The victims were also less likely to forgive the transgressors the more they took (both findings in line with expectations of participants of Study 1). Given the study's sample size, these results suggest that none of the strategies against the powerful transgressor work sufficiently well for their effect to be reliably detected.

In our sample, one-fifth of participants in a position of power (dictators) took everything from a victim regardless of how the victim reacted to their initial misbehaviour. These results are consistent with the literature, which shows that a fraction of people use purely antisocial strategies, maximizing only their own self-interest in social interactions or destroying the resources of others (Fehr and Charness 2023; Hart and Hare 1996; Kuběna et al. 2014). In an organization, this behaviour may manifest in various faces of destructive leadership (Schyns and Schilling 2013), like coercive power (Elangovan and Xie 2000), petty tyranny (Ashforth 1997) or tyrannical leadership (Hauge et al. 2007).

7 | Study 3

Study 3 investigates whether employees' real-world choices align with interpersonal dynamics predicted by restorative- and retributive-justice accounts. The conducted field survey enhances the ecological validity of previous studies by asking employees to report a recent experience of supervisor wrongdoing and to

describe how they responded, how the supervisor reacted, and what work-related consequences followed. We examine whether perceived severity is associated with a greater tendency to respond punitively, and whether punishment and forgiveness are differentially associated with supervisors' reconciliation and retaliation. We also evaluate whether response strategies are linked to the reported employees' subsequent work outcomes and well-being, and whether individual differences (empathy, political will and risk aversion) are associated with response selection. Full preregistered hypotheses, recruitment details, procedure, measures and analysis plan are reported in the [Supporting Information](#) section.

7.1 | Methods

7.1.1 | Participants

Using Prolific, we recruited a sample of 395 full-time employees (217 female, 178 male; $M_{\text{age}} = 37$ years, $SD_{\text{age}} = 9.95$) from the United Kingdom (74%) and the United States (26%). We invited participants using similar selection criteria as in Study 1. Because of the inability to reach the preregistered sample size of 400, we adjusted our sampling plan and included US participants as well. See [Supporting Information](#) section for details. Note that the change of the sampling plan may affect the interpretation of the reported p values (Simmons et al. 2011).

7.1.2 | Procedure and Measures

First, participants were asked whether they had experienced supervisor behaviour they perceived as unethical, unfair or excessively harsh towards them in the past 6 months and were willing to carefully read and honestly answer follow-up questions about that experience. Participants who did not confirm were screened out.

Those continuing to the main survey were asked to recall and briefly describe and categorize the most serious instance of such a supervisor's behaviour. Afterwards, they reported its perceived severity on a 7-point scale ranging from 'not severe at all' to 'extremely severe'. Then they categorized their own response to one of the three categories: punish/retaliate (e.g., reported the supervisor, formal complaint, collective action, refused extra work and sought sanctions), let it go/avoid (e.g., did nothing, minimized contact and waited it out) or forgive/repair (e.g., sought reconciliation and tried to empathize). In a similar way, they categorized the supervisor's reaction: retaliation (e.g., they harmed the participant), no reaction/no response (e.g., they carried on as usual) and guilt or reconciliation (e.g., they apologized or tried to repair their relationship with the participant). Participants also reported incident-related consequences regarding their workload, performance evaluation and monetary compensation; for each, they reported improvement, no change or worsening in the post-incident period, with an option to answer 'not sure/too early to tell'. In addition, they evaluated whether statements about post-incident experiences of increased work-related stress and burnout applied to them on a 7-point scale ranging from 'not at all' to 'very well'. Finally, participants filled the same measures of political will and empathy as in Study 2 as well as an incentivized risk aversion measure.

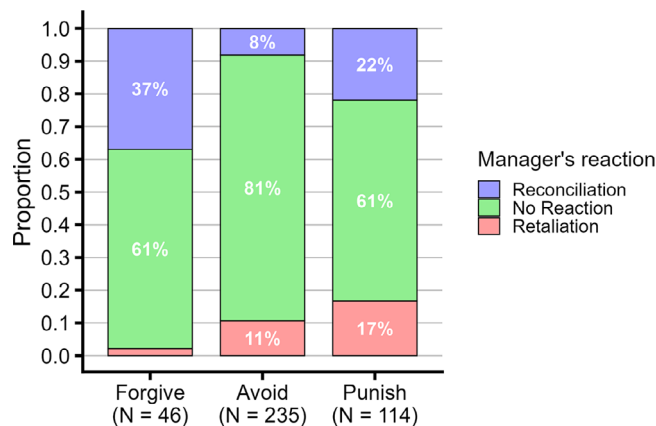


FIGURE 7 | Distribution of supervisors' reactions within each employee response category.

7.2 | Results

Participants most frequently reported avoiding/letting the incident go (59.5%), followed by punishing/retaliating (28.9%) and forgiving/repairing (11.6%). Supervisors were most often described as showing no reaction (73.2%), followed by guilt/reconciliation (15.4%) and retaliation (11.4%). A multinomial logistic regression predicting response category from perceived severity showed that higher severity predicted lower odds of avoid (vs. punish), Wald $z = -4.61$, $p < 0.001$, OR = 0.60, 95% CI [0.48, 0.74], and lower odds of forgive (vs. punish), Wald $z = -3.33$, $p < 0.001$, OR = 0.59, 95% CI [0.43, 0.80].

Among those who punished or forgave ($n = 160$), punishing (vs. forgiving) predicted higher odds of supervisor retaliation, Wald $z = 1.98$, $p = 0.048$, OR = 7.92, 95% CI [1.02, 61.61]. Punishing predicted lower odds of supervisor guilt/reconciliation, Wald $z = 1.96$, $p = 0.050$, OR = 0.46, 95% CI [0.21, 1.00]. Figure 7 depicts the conditional distribution of supervisors' reactions for each category of employee action.

Punishing (vs. forgiving) did not significantly predict reported workload change ($n = 145$), $t(140) = 0.391$, $p = 0.696$, $b = 0.143$, 95% CI [-0.58, 0.86], evaluation change ($n = 105$), $t(100) = 0.655$, $p = 0.512$, $b = 0.288$, 95% CI [-0.57, 1.15] or compensation change ($n = 141$), $t(136) = 1.44$, $p = 0.151$, $b = 0.822$, 95% CI [-0.30, 1.94]; punishing also did not predict incident-related stress ($n = 160$), $t(156) = -0.463$, $p = 0.644$, $b = -0.153$, 95% CI [-0.81, 0.50], or burnout ($n = 160$), $t(156) = 1.00$, $p = 0.319$, $b = 0.352$, 95% CI [-0.34, 1.05]. Finally, empathy did not predict forgiving (vs. avoid/punish), Wald $z = 0.532$, $p = 0.595$, OR = 1.20, 95% CI [0.61, 2.36]; and neither political will (Wald $z = 0.633$, $p = 0.527$, OR = 1.06, 95% CI [0.89, 1.25]) nor risk aversion (Wald $z = 1.516$, $p = 0.130$, OR = 1.15, 95% CI [0.96, 1.37]) predicted punishing (vs. avoid/forgive).

7.3 | Discussion

Study 3 extends the vignette and laboratory evidence by examining employees' reports of recent supervisor wrongdoing in the field. Avoidance was the most common response, occurring

substantially more frequently than punishment or forgiveness. This suggests that, in organizational power-imbalanced relationships, many targets prioritize self-protection over direct confrontation, consistent with the idea that perceived risk of retaliation and low efficacy constrain their actions. This overall pattern was broadly similar across different types of the supervisor's action as labelled by participants (see [Supporting Information Results](#) section). As in our other studies, perceived severity predicted a shift away from avoidance and forgiveness towards punishment. This pattern aligns with retributive intuitions, while also illustrating that escalation to punishment is most likely when transgressions are perceived as serious enough to justify the interpersonal and career risks associated with challenging them.

We also replicated the finding that punishment was associated with a higher likelihood of supervisor retaliation and a lower likelihood of reconciliation relative to forgiveness. This field pattern mirrors the backlash found in Studies 1 and 2a/b, supporting the conclusion that punitive responses against a powerful transgressor can be strategically fragile. Nevertheless, we did not detect reliable associations between punishing (vs. forgiving) and downstream changes in workload, evaluations, compensation or well-being. One interpretation is that these outcomes depend heavily on the organizational context (e.g., HR procedures, team support and labour protections) and thus may not be well captured in a retrospective, single-incident design. Another is that any benefits (or costs) of the action unfold over longer horizons, as evidenced by the fact that many participants reported uncertainty about consequences.

Finally, Study 3 replicates weak or null associations between response strategy and the individual-difference measures examined. This null pattern suggests that, in the field, situational constraints may dominate stable traits in shaping whether targets punish, avoid or forgive.

8 | General Discussion

Across studies, the severity of supervisors' misbehaviour reliably shaped subordinates' responses. In Study 1, participants expected more punishment for severe infractions and more forgiveness for minor ones and anticipated harsher reactions to punishment and kinder reactions to forgiveness, particularly when the infraction was less severe. Studies 2a/2b replicated the severity–response link: greater withdrawals increased punishment and reduced forgiveness. Yet these responses did not improve power holders' behaviour; dictators were largely unresponsive and, if anything, became more unethical. Study 3 showed similar dynamics in the field as employees most often let incidents go; higher perceived severity predicted punishment, which in turn was linked to more retaliation and less reconciliation, without reliable downstream associations with workload, evaluations, compensation, stress or burnout. Overall, the four studies reveal an expectation–outcome gap, i.e., punishment and forgiveness are salient responses, but they do not reliably produce moral repair when transgressors retain structural advantage.

The lack of responsiveness from superiors to subordinates' punitive or forgiving actions in Studies 2a/b and 3 suggests

that power holders may feel insulated from the consequences of their unethical behaviour. This aligns with theories suggesting that power can reduce empathy and concern for others, leading to a decreased likelihood of altering behaviour in response to subordinates' actions (Foulek et al. 2018, 2020; Galinsky et al. 2006). Consistent with this account, Study 3 indicates that punitive responses are more often met with retaliation than with reconciliation, suggesting that power not only reduces responsiveness to subordinates' signals but can also convert confrontation into additional harm for the target.

Related evidence comes from research on interpersonal emotion communication in negotiation, which examines how people strategically display anger, disappointment or happiness to shape counterparts' inferences and behaviour (Ye et al. 2023, 2025; van Dijk et al. 2008, 2018). In this literature, communicating anger towards an unfair offer functions similarly to a punitive signal, conveying disapproval and high limits (van Dijk et al. 2008), whereas communicating happiness or warmth (especially while experiencing anger) functions similarly to a conciliatory or relationship-maintenance signal (Ye et al. 2023). Importantly, the effectiveness of these signals is power-contingent as communicated anger can elicit concessions when it credibly signals toughness, but it can also backfire when the target has a structural advantage. Moreover, when dealing with a high-power counterpart, negotiators tend to strategically shift from communicating anger to communicating disappointment, a less confrontational signal that may reduce backlash risk while still conveying discontent (van Dijk et al. 2018). These findings fit our expectation–outcome gap as targets anticipate relational and behavioural consequences from punitive versus conciliatory responses, yet power holders may remain unresponsive or retaliate when accountability is low.

Several design features warrant caution. First, Studies 1 and 2a used restricted choice menus, producing conditional (pairwise) choices rather than an unconstrained three-way ranking of punishment, ignoring and forgiveness; we therefore treat the results as menu-dependent tendencies, not absolute preferences. Second, the laboratory interaction had a known finite horizon (two rounds), and end-round effects are common in economic games (Andreoni 1988). In a terminal round, a payoff-maximizing dictator has little incentive to moderate taking because there is no future interaction in which the target can reciprocate. Thus, even if punishment or forgiveness changes the target's beliefs or the dictator's momentary affect, dictators who were already pursuing a payoff-maximizing strategy in Round 1 may have little reason to adjust in Round 2 when the game is known to end. Accordingly, the null effects of punishment and forgiveness on Round-2 withdrawals should be interpreted as evidence of limited short-horizon malleability rather than as a test of moral repair processes in ongoing hierarchical relationships (Gordon and Puurtinen 2025). At the same time, because the finite horizon was identical across conditions, our main identification still concerns between-condition differences under the same end-game incentives. Third, Study 1 relied on hypothetical vignettes that may reflect social desirability or reputational motives and may not fully translate into incentivized behaviour (Vranka and Houdek 2024). Study 3 is retrospective and correlational, preventing causal claims, and relies on single-incident self-reports that may be shaped by recall and motivated interpretation.

More broadly, the laboratory setting in Studies 2a/b is artificial and short-term relative to real organizational wrongdoing (Galizzi and Navarro-Martínez 2019; Levitt and List 2007): Participants may have lacked the investment or fear of repercussions that employees face, multi-round interactions still compress longer time horizons and anonymity precluded reputational dynamics. Finally, cultural variation across the USA, UK, Poland, and the Czech context may limit generalizability to other settings (Taras et al. 2010).

9 | Conclusion and Future Research

We showed that subordinates want to retaliate when a superior's abuse is severe, yet that retaliation neither reforms the superior nor improves pay-offs. Even among transgressors higher in guilt or shame-proneness, forgiveness did not reliably predict corrective action, and punishment actually increased retaliation among transgressors who were higher in shame-proneness. Our research makes several contributions to the organizational and social psychology literature on victims' responses to unethical actions by powerful transgressors.

First, according to retributive theories (Fehr and Gächter 2002; Greve and Teh 2016; Rupp and Bell 2010), individuals are willing to penalize unethical behaviour, even at a personal cost, without direct material benefit. However, this approach proves ineffective in isolated dyadic interactions where power imbalances exist, as the enforcement of justice requires the collective action of individuals. Punishment in unequal games is often fuelled by anger rather than strategic deterrence; angry acts can be reflexive and hence ineffectual (Galeotti 2015; Gneezy and Imas 2014).

When the transgressor is powerful, punishment tends to backfire. More severe wrongdoing elicited stronger subordinate punishment, which then triggered even harsher retaliation. Consistent with bargaining research (Khan and Howe 2021), anger-based punishment is ineffective against relatively unconstrained targets, and organizational punishment likely differs in threat level (confrontational moral anger compared to low-threat disapproval), which future work should separate when assessing retaliation risk. Forgiveness and guilt or shame induction also failed to secure justice and harmed empathic subordinates most, paralleling the 'Justine effect', where altruists disproportionately attract punishment from free riders (Kuběna et al. 2014; Pleasant and Barclay 2018).

It is understandable that many participants refrained from punishing unethical behaviour by transgressors, particularly when their choice was limited to ignoring or punishing the misconduct. Although some researchers describe such inaction as complicity that enables wrongdoing to persist (Knoll and van Dick 2013; MacGregor and Stuebs 2014), our results suggest it may be a rational response shaped by experience. Targets' (in)action is governed by perceived efficacy and the availability of identity-based support. The social identity model of collective action (SIMCA) emphasizes perceived efficacy as a key proximal driver of action, alongside perceptions of injustice and identification (van Zomeren et al. 2008). In hierarchical settings, when individuals anticipate low efficacy and high retaliation risk, avoidance or

silence can be instrumentally rational even when punishment is normatively endorsed.

These outcomes may result from social norms that dictate that transgressors should be fought at all costs or forgiven at a high cost, even when, as in our experiments, they have absolute unchecked power and in no way reflect the behaviour of the victims. On the other hand, negotiation research suggests that positive emotion communication can function as an explicit relational investment signal even when the communicator privately feels anger about unfairness. This work supports the more general idea that kindness displays may be strategically deployed to stabilize hierarchical relationships and secure downstream benefits, even when they do not reform the powerful actor's underlying preferences.

Second, the persistence of antisocial behaviour among some powerful transgressors highlights how difficult ethical violations are to manage, as moral repair theory anticipates. Such actors may resist disciplinary, corrective or reconciliatory responses, and the theory offers limited leverage because it does not clearly distinguish genuine rebuilding of trust from performative signalling that may be easier when the offender is weak. When transgressors hold greater power, community disapproval, anger, or sanctions may have less deterrent force, and guilt or shame may be absent, increasing the likelihood that offenders ignore restorative efforts or even escalate misconduct (Houdek et al. 2021; Cortina 2008; Cortina and Magley 2003). Given limited tools to diagnose sincere restoration, moral repair theory may provide insufficient guidance for sanctioning offenders who can evade accountability (Vives-Gabriel et al. 2024), highlighting the need for stronger anti-abuse cultures (Sutton 2007) and regulatory oversight.

The literature on third-party punishment may be informative in these dilemmas. Uninvolved observers often prefer compensation of the victim to punishment of the perpetrator when the two are mutually exclusive (Chavez and Bicchieri 2013; Fehr and Fischbacher 2004; Fehr and Charness 2023). Our design offered subordinates only three first-party moves. Adding a cost-equivalent compensate-me option (or a reputational report to a neutral arbiter) may reveal paths to repair when direct deterrence fails.

Acknowledgements

For valuable comments and advice on earlier versions of the manuscript, we thank participants at various workshops and discussion seminars, most notably EURAM 2024, IAREP/SABE 2024, and SJDM 2024. We would like to thank Richard Kaucký, Kristína Klovaničová, Jan Kolín, Nicolas Say, Lenka Valerianová and Natálie Zogno for their help with data collection for Study 2a/b. Thanks also go to Marek Hudík for an insightful discussion of the original idea. We would also like to thank the editor and two anonymous reviewers for their highly detailed comments and suggestions, which enriched the final article.

Funding

The work on this article was supported by the Czech Science Agency (GACR) project No. 22-29520S, 'Behavioral Organizational Politics: Experiments in Prosocial Political Behavior'.

Ethics Statement

The study was performed in accordance with the principles of the Declaration of Helsinki and all relevant regulations for conducting psychological studies in the Czech Republic.

Consent

All participants provided informed consent and were assured of anonymity and confidentiality of their responses.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that supports the findings of this study are available in the [Supporting Information](#) section of this article.

Endnotes

¹Data for Study 2a were collected first, but the studies are described in this order for the sake of better exposition of the arguments by providing real-world examples for the situations modelled in the economic game used in Study 2.

²There were no substantial differences in the directions of the effects for the respondents of the two nationalities, only some differences in their sizes. We, therefore, report the analyses with the data from participants of the two nationalities combined.

³We did not have specific hypotheses regarding the expectations, so the results of the analysis of dictators' expectations are reported in the Supplementary results.

References

- Amore, M. D., O. Garofalo, and A. Guerra. 2023. "How Leaders Influence (un)Ethical Behaviors Within Organizations: A Laboratory Experiment on Reporting Choices." *Journal of Business Ethics* 183, no. 2: 495–510. <https://doi.org/10.1007/s10551-022-05088-z>.
- Andreoni, J. 1988. "Why Free Ride?: Strategies and Learning in Public Goods Experiments." *Journal of Public Economics* 37, no. 3: 291–304. [https://doi.org/10.1016/0047-2727\(88\)90043-6](https://doi.org/10.1016/0047-2727(88)90043-6).
- Ashforth, B. E. 1997. "Petty Tyranny in Organizations: A Preliminary Examination of Antecedents and Consequences." *Canadian Journal of Administrative Sciences/Revue Canadienne Des Sciences de l'Administration* 14, no. 2: 126–140. <https://doi.org/10.1111/j.1936-4490.1997.tb00124.x>.
- Bandura, A. 1999. "Moral Disengagement in the Perpetration of Inhumanities." *Personality and Social Psychology Review* 3, no. 3: 193–209. https://doi.org/10.1207/s15327957pspr0303_3.
- Berndsen, M., and M. Wenzel. 2021. "Offenders' Claims of Taking the Victims' Perspective Can Promote Forgiveness, or Backfire! the Moderating Role of Correctly Voicing the Victims' Emotions in Collective Apologies." *European Journal of Social Psychology* 51, no. 1: 5–22. <https://doi.org/10.1002/ejsp.2710>.
- Bloom, P. 2017. *Against Empathy: The Case for Rational Compassion*. Random House.
- Bosman, R., M. Sutter, and F. van Winden. 2005. "The Impact of Real Effort and Emotions in the Power-to-Take Game." *Journal of Economic Psychology* 26, no. 3: 407–429. <https://doi.org/10.1016/j.joep.2004.12.005>.
- Bosman, R., and F. Van Winden. 2002. "Emotional Hazard in a Power-to-Take Experiment." *Economic Journal* 112, no. 476: 147–169. <https://doi.org/10.1111/1468-0297.0j677>.
- Bottom, W. P., K. Gibson, S. E. Daniels, and J. K. Murnighan. 2002. "When Talk Is Not Cheap: Substantive Penance and Expressions of Intent in

Rebuilding Cooperation." *Organization Science* 13, no. 5: 497–513. <https://doi.org/10.1287/orsc.13.5.497.7816>.

Brandt, H., C. Hauert, and K. Sigmund. 2006. "Punishing and Abstaining for Public Goods." *Proceedings of the National Academy of Sciences of the United States of America* 103, no. 2: 495–497. <https://doi.org/10.1073/pnas.0507229103>.

Burris, E. R. 2012. "The Risks and Rewards of Speaking Up: Managerial Responses to Employee Voice." *Academy of Management Journal* 55, no. 4: 851–875. <https://doi.org/10.5465/amj.2010.0562>.

Butterfield, K. D., and J. Goodstein. 2010. "Extending the Horizon of Business Ethics: Restorative Justice and the Aftermath of Unethical Behavior." *Business Ethics Quarterly* 20, no. 3: 453–480. <https://doi.org/10.5840/beq201020330>.

Butterfield, K. D., N. R. Neale, E. Shin, and M. He (Rebecca). 2023. "Moral Repair Versus Punishment: Influences on Managerial Responses." *Organization Management Journal* 20, no. 4: 169–180. <https://doi.org/10.1108/OMJ-11-2021-1398>.

Chavez, A. K., and C. Bicchieri. 2013. "Third-Party Sanctioning and Compensation Behavior: Findings From the Ultimatum Game." *Journal of Economic Psychology* 39: 268–277. <https://doi.org/10.1016/j.joep.2013.09.004>.

Chiang, J. T.-J., H. Liu, R. Fehr, Z. Wang, and Q. Huang. 2023. "Leaders and the Punishment of Misconduct: Examining the Roles of Leader Moral Identity and Cognitive Load." *Journal of Applied Psychology* 109: 1022–1038. <https://doi.org/10.1037/apl0001108>.

Cohen, T. R., A. T. Panter, and N. Turan. 2012. "Guilt Proneness and Moral Character." *Current Directions in Psychological Science* 21, no. 5: 355–359. <https://doi.org/10.1177/0963721412454874>.

Cohen, T. R., A. T. Panter, and N. Turan. 2013. "Predicting Counterproductive Work Behavior From Guilt Proneness." *Journal of Business Ethics* 114, no. 1: 45–53. <https://doi.org/10.1007/s10551-012-1326-2>.

Cortina, L. M. 2008. "Unseen Injustice: Incivility as Modern Discrimination in Organizations." *Academy of Management Review* 33, no. 1: 55–75. <https://doi.org/10.5465/amr.2008.27745097>.

Cortina, L. M., and V. J. Magley. 2003. "Raising Voice, Risking Retaliation: Events Following Interpersonal Mistreatment in the Workplace." *Journal of Occupational Health Psychology* 8, no. 4: 247–265. <https://doi.org/10.1037/1076-8998.8.4.247>.

Crawford, J. R., and J. D. Henry. 2004. "The Positive and Negative Affect Schedule (PANAS): Construct Validity, Measurement Properties and Normative Data in a Large Non-Clinical Sample." *British Journal of Clinical Psychology* 43, no. 3: 245–265. <https://doi.org/10.1348/0144665031752934>.

Deshpande, S. P., E. George, and J. Joseph. 2000. "Ethical Climates and Managerial Success in Russian Organizations." *Journal of Business Ethics* 23, no. 2: 211–217. <https://doi.org/10.1023/a:1005943017693>.

Detert, J. R., and E. R. Burris. 2007. "Leadership Behavior and Employee Voice: Is the Door Really Open?" *Academy of Management Journal* 50, no. 4: 869–884. <https://doi.org/10.5465/amj.2007.26279183>.

Doliński, D. 2018. "Is Psychology Still a Science of Behaviour?" *Social Psychological Bulletin* 13, no. 2: 1–14. <https://doi.org/10.5964/spb.v13i2.25025>.

Dreber, A., D. G. Rand, D. Fudenberg, and M. A. Nowak. 2008. "Winners Don't Punish." *Nature* 452, no. 7185: 348–351. <https://doi.org/10.1038/nature06723>.

Duersch, P., and J. Müller. 2015. "Taking Punishment Into Your Own Hands: An Experiment." *Journal of Economic Psychology* 46: 1–11. <https://doi.org/10.1016/j.joep.2014.10.004>.

Dungan, J. A., L. Young, and A. Waytz. 2019. "The Power of Moral Concerns in Predicting Whistleblowing Decisions." *Journal of Experimental Social Psychology* 85: 103848. <https://doi.org/10.1016/j.jesp.2019.103848>.

Elangovan, A. R., and J. L. Xie. 2000. "Effects of Perceived Power of Supervisor on Subordinate Work Attitudes." *Leadership &*

- Organization Development Journal* 21, no. 6: 319–328. <https://doi.org/10.1108/01437730010343095>.
- Engel, C. 2011. “Dictator Games: A Meta Study.” *Experimental Economics* 14, no. 4: 583–610. <https://doi.org/10.1007/s10683-011-9283-7>.
- Enright, R. D., S. Freedman, and J. Rique. 1998. “The Psychology of Interpersonal Forgiveness.” In *Exploring Forgiveness*, edited by R. D. Enright and J. North, 46–62. University of Wisconsin Press.
- Ent, M. R., and R. F. Baumeister. 2015. “Individual Differences in Guilt Proneness Affect How People Respond to Moral Tradeoffs Between Harm Avoidance and Obedience to Authority.” *Personality and Individual Differences* 74: 231–234. <https://doi.org/10.1016/j.paid.2014.10.035>.
- Erlandsson, A., F. Björklund, and M. Bäckström. 2015. “Emotional Reactions, Perceived Impact and Perceived Responsibility Mediate the Identifiable Victim Effect, Proportion Dominance Effect and In-Group Effect Respectively.” *Organizational Behavior and Human Decision Processes* 127: 1–14. <https://doi.org/10.1016/j.obhdp.2014.11.003>.
- Exline, J. J., R. F. Baumeister, A. L. Zell, A. J. Kraft, and C. V. O. Witvliet. 2008. “Not So Innocent: Does Seeing One’s Own Capability for Wrongdoing Predict Forgiveness?” *Journal of Personality and Social Psychology* 94, no. 3: 495–515. <https://doi.org/10.1037/0022-3514.94.3.495>.
- Fatas, E., and L. Restrepo-Plaza. 2022. “When Losses Can be a Gain. A Large Lab-in-the-Field Experiment on Reference Dependent Forgiveness in Colombia.” *Journal of Economic Psychology* 88: 102463. <https://doi.org/10.1016/j.joep.2021.102463>.
- Fehr, E., and G. Charness. 2023. “Social Preferences: Fundamental Characteristics and Economic Consequences.” *Journal of Economic Literature* 63: 440–514. <https://doi.org/10.1257/jel.20241391>.
- Fehr, E., and U. Fischbacher. 2004. “Third-party Punishment and Social Norms.” *Evolution and Human Behavior* 25, no. 2: 63–87. [https://doi.org/10.1016/S1090-5138\(04\)00005-4](https://doi.org/10.1016/S1090-5138(04)00005-4).
- Fehr, E., and S. Gächter. 2000. “Cooperation and Punishment in Public Goods Experiments.” *American Economic Review* 90, no. 4: 980–994.
- Fehr, E., and S. Gächter. 2002. “Altruistic Punishment in Humans.” *Nature* 415, no. 6868: 137–140. <https://doi.org/10.1038/415137a>.
- Foulk, T. A., N. Chighizola, and G. Chen. 2020. “Power Corrupts (or Does It?): An Examination of the Boundary Conditions of the Antisocial Effects of Experienced Power.” *Social and Personality Psychology Compass* 14, no. 4: e12524. <https://doi.org/10.1111/spc3.12524>.
- Foulk, T. A., K. Lanaj, M.-H. Tu, A. Erez, and L. Archambeau. 2018. “Heavy Is the Head That Wears the Crown: An Actor-Centric Approach to Daily Psychological Power, Abusive Leader Behavior, and Perceived Incivility.” *Academy of Management Journal* 61, no. 2: 661–684. <https://doi.org/10.5465/amj.2015.1061>.
- Frieder, R. E., G. R. Ferris, P. L. Perrewé, A. Wihler, and C. D. Brooks. 2019. “Extending the Metatheoretical Framework of Social/Political Influence to Leadership: Political Skill Effects on Situational Appraisals, Responses, and Evaluations by Others.” *Personnel Psychology* 72, no. 4: 543–569. <https://doi.org/10.1111/peps.12336>.
- Frollová, N., M. Vranka, and P. Houdek. 2021. “A Qualitative Study of Perception of a Dishonesty Experiment.” *Journal of Economic Methodology* 28, no. 3: 274–290. <https://doi.org/10.1080/1350178X.2021.1936598>.
- Fudenberg, D., D. G. Rand, and A. Dreber. 2012. “Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World.” *American Economic Review* 102, no. 2: 720–749. <https://doi.org/10.1257/aer.102.2.720>.
- Gago, A. 2021. “Reciprocity and Uncertainty: When Do People Forgive?” *Journal of Economic Psychology* 84: 102362. <https://doi.org/10.1016/j.joep.2021.102362>.
- Galeotti, F. 2015. “Do Negative Emotions Explain Punishment in Power-to-Take Game Experiments?” *Journal of Economic Psychology* 49: 1–14. <https://doi.org/10.1016/j.joep.2015.03.005>.
- Galinsky, A. D., J. C. Magee, M. Ena Inesi, and D. H. Gruenfeld. 2006. “Power and Perspectives Not Taken.” *Psychological Science* 17, no. 12: 1068–1074. <https://doi.org/10.1111/j.1467-9280.2006.01824.x>.
- Galizzi, M. M., and D. Navarro-Martinez. 2019. “On the External Validity of Social Preference Games: A Systematic Lab-Field Study.” *Management Science* 65, no. 3: 976–1002. <https://doi.org/10.1287/mnsc.2017.2908>.
- Gneezy, U., and A. Imas. 2014. “Materazzi Effect and the Strategic Use of Anger in Competitive Interactions.” *Proceedings of the National Academy of Sciences* 111, no. 4: 1334–1337. <https://doi.org/10.1073/pnas.1313789111>.
- Goodstein, J., K. D. Butterfield, and N. Neale. 2016. “Moral Repair in the Workplace: A Qualitative Investigation and Inductive Model.” *Journal of Business Ethics* 138, no. 1: 17–37. <https://doi.org/10.1007/s10551-015-2593-5>.
- Gordon, D., and M. Puurtinen. 2025. “Fairness Is What You Can Get Away With: Proposer and Responder Behaviour in a Collective Action Ultimatum Game.” *Social Psychology Bulletin* 20: 1–33. <https://doi.org/10.32872/spb.11607>.
- Graham, J., B. A. Nosek, J. Haidt, R. Iyer, S. Koleva, and P. H. Ditto. 2011. “Mapping the Moral Domain.” *Journal of Personality and Social Psychology* 101, no. 2: 366–385. <https://doi.org/10.1037/a0021847>.
- Greve, H. R., D. Palmer, and J. Pozner. 2010. “Organizations Gone Wild: The Causes, Processes, and Consequences of Organizational Misconduct.” *Academy of Management Annals* 4, no. 1: 53–107. <https://doi.org/10.1080/19416521003654186>.
- Greve, H. R., and D. the. 2016. “Consequences of Organizational Misconduct: Too Much and Too Little Punishment.” In *Organizational Wrongdoing: Key Perspectives and New Directions*, edited by R. Greenwood, D. Palmer, and K. Smith-Crowe, 370–403. Cambridge University Press. <https://doi.org/10.1017/CBO9781316338827.014>.
- Hareli, S., and Z. Eisikovits. 2006. “The Role of Communicating Social Emotions Accompanying Apologies in Forgiveness.” *Motivation and Emotion* 30, no. 3: 189–197. <https://doi.org/10.1007/s11031-006-9025-x>.
- Hart, S., and R. Hare. 1996. “Psychopathy and Antisocial Personality Disorder.” *Current Opinion in Psychiatry* 9, no. 2: 129–132.
- Hauge, J. L., A. Skogstad, and S. Einarsen. 2007. “Relationships Between Stressful Work Environments and Bullying: Results of a Large Representative Study.” *Work & Stress* 21, no. 3: 220–242. <https://doi.org/10.1080/02678370701705810>.
- Heffner, J., and O. FeldmanHall. 2019. “Why We Don’t Always Punish: Preferences for Non-Punitive Responses to Moral Violations.” *Scientific Reports* 9, no. 1: 13219. <https://doi.org/10.1038/s41598-019-49680-2>.
- Holt, C. A., and S. K. Laury. 2002. “Risk Aversion and Incentive Effects.” *American Economic Review* 92, no. 5: 1644–1655. <https://doi.org/10.1257/000282802762024700>.
- Houdek, P., Š. Bahník, M. Hudík, and M. Vranka. 2021. “Selection Effects on Dishonest Behavior.” *Judgment & Decision Making* 16, no. 2: 238–266. <https://doi.org/10.1017/S1930297500008561>.
- Hu, Y., S. Qiu, G. Wang, et al. 2024. “Are Guilt-Prone Power-Holders Less Corrupt? Evidence From Two Online Experiments.” *Social Psychological and Personality Science* 15, no. 4: 430–438. <https://doi.org/10.1177/19485506231168515>.
- Jadaszewski, S., S. L. Speight, N. Alshabani, et al. 2024. ““It Felt Like Such a Closed Door”: Supervisory Cultural Rupture & Humility.” *Counseling Psychologist* 52, no. 2: 298–338. <https://doi.org/10.1177/00110000231217070>.
- Jordan, J. J., M. Hoffman, P. Bloom, and D. G. Rand. 2016. “Third-party Punishment as a Costly Signal of Trustworthiness.” *Nature* 530, no. 7591: 473–476. <https://doi.org/10.1038/nature16981>.
- Kakkar, H., N. Sivanathan, and M. S. Gobel. 2020. “Fall From Grace: The Role of Dominance and Prestige in the Punishment of High-Status Actors.” *Academy of Management Journal* 63, no. 2: 530–553. <https://doi.org/10.5465/amj.2017.0729>.

- Kapoutsis, I., A. Papalexandris, D. C. Treadway, and J. Bentley. 2017. "Measuring Political Will in Organizations: Theoretical Construct Development and Empirical Validation." *Journal of Management* 43, no. 7: 2252–2280. <https://doi.org/10.1177/0149206314566460>.
- Kazarovytzka, F., and R. Imhoff. 2022. "Too Great to be Guilty? Individuals High in Collective Narcissism Demand Closure Regarding the Past to Attenuate Collective Guilt." *European Journal of Social Psychology* 52, no. 4: 748–771. <https://doi.org/10.1002/ejsp.2850>.
- Khan, S. R., and L. C. Howe. 2021. "Concern for the Transgressor's Consequences: An Explanation for Why Wrongdoing Remains Unreported." *Journal of Business Ethics* 173, no. 2: 325–344. <https://doi.org/10.1007/s10551-020-04568-4>.
- Kiewitz, C., S. L. D. Restubog, M. K. Shoss, P. R. J. M. Garcia, and R. L. Tang. 2016. "Suffering in Silence: Investigating the Role of Fear in the Relationship Between Abusive Supervision and Defensive Silence." *Journal of Applied Psychology* 101, no. 5: 731–742. <https://doi.org/10.1037/apl0000074>.
- Klimecki, O. M. 2019. "The Role of Empathy and Compassion in Conflict Resolution." *Emotion Review* 11, no. 4: 310–325. <https://doi.org/10.1177/1754073919838609>.
- Knoll, M., and R. van Dick. 2013. "Do I Hear the Whistle...? A First Attempt to Measure Four Forms of Employee Silence and Their Correlates." *Journal of Business Ethics* 113, no. 2: 349–362. <https://doi.org/10.1007/s10551-012-1308-4>.
- Kuběna, A. A., P. Houdek, J. Lindová, L. Připlová, and J. Flegr. 2014. "Justine Effect: Punishment of the Unduly Self-Sacrificing Cooperative Individuals." *PLoS ONE* 9, no. 3: e92336. <https://doi.org/10.1371/journal.pone.0092336>.
- Kurzban, R., M. E. McCullough, and B. A. Tabak. 2013. "Cognitive Systems for Revenge and Forgiveness." *Behavioral and Brain Sciences* 36, no. 1: 1–15. <https://doi.org/10.1017/S0140525X11002160>.
- LaVan, H., and W. M. Martin. 2008. "Bullying in the U.S. Workplace: Normative and Process-Oriented Ethical Approaches." *Journal of Business Ethics* 83, no. 2: 147–165. <https://doi.org/10.1007/s10551-007-9608-9>.
- Leana, C. R., V. Mittal, and E. Stiehl. 2012. "Organizational Behavior and the Working Poor." *Organization Science* 23, no. 3: 888–906. <https://doi.org/10.1287/orsc.1110.0672>.
- Lerner, J. S., and D. Keltner. 2001. "Fear, Anger, and Risk." *Journal of Personality and Social Psychology* 81, no. 1: 146–159. <https://doi.org/10.1037/0022-3514.81.1.146>.
- Levitt, S. D., and J. A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21, no. 2: 153–174. <https://doi.org/10.1257/jep.21.2.153>.
- List, J. A. 2007. "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy* 115, no. 3: 482–493. <https://doi.org/10.1086/519249>.
- Liu, Y., S. Zhao, R. Li, L. Zhou, and F. Tian. 2018. "The Relationship Between Organizational Identification and Internal Whistle-Blowing: The Joint Moderating Effects of Perceived Ethical Climate and Proactive Personality." *Review of Managerial Science* 12, no. 1: 113–134. <https://doi.org/10.1007/s11846-016-0214-z>.
- MacGregor, J., and M. Stuebs. 2014. "The Silent Samaritan Syndrome: Why the Whistle Remains Unblown." *Journal of Business Ethics* 120, no. 2: 149–164. <https://doi.org/10.1007/s10551-013-1639-9>.
- McCullough, M. E. 2001. "Forgiveness: Who Does It and How Do They Do It?" *Current Directions in Psychological Science* 10, no. 6: 194–197. <https://doi.org/10.1111/1467-8721.00147>.
- Miceli, M. P., and J. P. Near. 1994. "Whistleblowing: Reaping the Benefits." *Academy of Management Perspectives* 8, no. 3: 65–72. <https://doi.org/10.5465/ame.1994.9503101177>.
- Miceli, M. P., and J. P. Near. 2005. "Standing Up or Standing By: What Predicts Blowing the Whistle on Organizational Wrongdoing?" In *Research in Personnel and Human Resources Management*, edited by J. J. Martocchio, 95–136. Emerald Group Publishing Limited. [https://doi.org/10.1016/S0742-7301\(05\)24003-3](https://doi.org/10.1016/S0742-7301(05)24003-3).
- Moore, C. 2015. "Moral Disengagement." *Current Opinion in Psychology* 6: 199–204. <https://doi.org/10.1016/j.copsyc.2015.07.018>.
- Morrison, E. W. 2014. "Employee Voice and Silence." *Annual Review of Organizational Psychology and Organizational Behavior* 1, no. 1: 173–197. <https://doi.org/10.1146/annurev-orgpsych-031413-091328>.
- Nikiforakis, N. 2008. "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" *Journal of Public Economics* 92, no. 1: 91–112. <https://doi.org/10.1016/j.jpubeco.2007.04.008>.
- Nikiforakis, N., H.-T. Normann, and B. Wallace. 2010. "Asymmetric Enforcement of Cooperation in a Social Dilemma." *Southern Economic Journal* 76, no. 3: 638–659. <https://doi.org/10.4284/sej.2010.76.3.638>.
- Nikiforakis, N., C. N. Noussair, and T. Wilkening. 2012. "Normative Conflict and Feuds: The Limits of Self-Enforcement." *Journal of Public Economics* 96, no. 9: 797–807. <https://doi.org/10.1016/j.jpubeco.2012.05.014>.
- Nockur, L., S. Pfattheicher, and J. Keller. 2021. "Different Punishment Systems in a Public Goods Game With Asymmetric Endowments." *Journal of Experimental Social Psychology* 93: 104096. <https://doi.org/10.1016/j.jesp.2020.104096>.
- Norlock, K. J. 2022. "Forgiveness and Moral Repair." In *The Oxford Handbook of Moral Psychology*, edited by M. Vargas and J. M. Doris, 929–946. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198871712.013.46>.
- Novaro, R., G. Nasi, M. Cucciniello, and S. Grimmelikhuisen. 2024. "The Power of Framing: The Role of Information Provision in Promoting Whistleblowing." *Public Administration* 102, no. 4: 1342–1365. <https://doi.org/10.1111/padm.12977>.
- Oelrich, S. 2021. "Intention Without Action? Differences Between Whistleblowing Intention and Behavior on Corruption and Fraud." *Business Ethics, the Environment & Responsibility* 30, no. 3: 447–463. <https://doi.org/10.1111/beer.12337>.
- Pleasant, A., and P. Barclay. 2018. "Why Hate the Good Guy? Antisocial Punishment of High Cooperators Is Greater When People Compete to be Chosen." *Psychological Science* 29, no. 6: 868–876. <https://doi.org/10.1177/0956797617752642>.
- Riek, B. M., and E. W. Mania. 2012. "The Antecedents and Consequences of Interpersonal Forgiveness: A Meta-Analytic Review." *Personal Relationships* 19, no. 2: 304–325. <https://doi.org/10.1111/j.1475-6811.2011.01363.x>.
- Robinson, S. L., and R. J. Bennett. 1995. "A Typology of Deviant Workplace Behaviors: A Multidimensional Scaling Study." *Academy of Management Journal* 38, no. 2: 555–572. <https://doi.org/10.5465/256693>.
- Rupp, D. E., and C. M. Bell. 2010. "Extending the Deontic Model of Justice: Moral Self-Regulation in Third-Party Responses to Injustice." *Business Ethics Quarterly* 20, no. 1: 89–106. <https://doi.org/10.5840/beq20102017>.
- Schyns, B., and J. Schilling. 2013. "How Bad Are the Effects of Bad Leaders? A Meta-Analysis of Destructive Leadership and Its Outcomes." *Leadership Quarterly* 24, no. 1: 138–158. <https://doi.org/10.1016/j.leaqua.2012.09.001>.
- Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22, no. 11: 1359–1366. <https://doi.org/10.1177/0956797611417632>.
- Spreng, R. N., M. C. McKinnon, R. A. Mar, and B. Levine. 2009. "The Toronto Empathy Questionnaire: Scale Development and Initial Validation of a Factor-Analytic Solution to Multiple Empathy Measures." *Journal of Personality Assessment* 91, no. 1: 62–71. <https://doi.org/10.1080/00223890802484381>.
- Sutton, R. I. 2007. *The No Asshole Rule: Building a Civilized Workplace and Surviving One That Isn't*. Business Plus.

- Tangney, J. P. 1995. "Shame and Guilt in Interpersonal Relationships." In *Self-Conscious Emotions: The Psychology of Shame, Guilt, Embarrassment, and Pride*. Guilford Press.
- Tangney, J. P., and R. L. Dearing. 2002. *Shame and Guilt*. Guilford Press.
- Tangney, J. P., J. Stuewig, and D. J. Mashek. 2007. "Moral Emotions and Moral Behavior." *Annual Review of Psychology* 58: 345–372. <https://doi.org/10.1146/annurev.psych.56.091103.070145>.
- Taras, V., B. L. Kirkman, and P. Steel. 2010. "Examining the Impact of Culture's Consequences: A Three-Decade, Multilevel, Meta-Analytic Review of Hofstede's Cultural Value Dimensions." *Journal of Applied Psychology* 95, no. 3: 405–439. <https://doi.org/10.1037/a0018938>.
- Teubner, T., M. Adam, and C. Niemeyer. 2015. "Measuring Risk Preferences in Field Experiments: Proposition of a Simplified Task." *Economics Bulletin* 35, no. 3: 1510–1517.
- van Dijk, E., I. van Beest, G. A. van Kleef, and G.-J. Lelieveld. 2018. "Communication of Anger Versus Disappointment in Bargaining and the Moderating Role of Power." *Journal of Behavioral Decision Making* 31, no. 5: 632–643. <https://doi.org/10.1002/bdm.2079>.
- van Dijk, E., G. A. van Kleef, W. Steinel, and I. van Beest. 2008. "A Social Functional Approach to Emotions in Bargaining: When Communicating Anger Pays and When It Backfires." *Journal of Personality and Social Psychology* 94, no. 4: 600–614. <https://doi.org/10.1037/0022-3514.94.4.600>.
- van Zomeren, M., T. Postmes, and R. Spears. 2008. "Toward an Integrative Social Identity Model of Collective Action: A Quantitative Research Synthesis of Three Socio-Psychological Perspectives." *Psychological Bulletin* 134, no. 4: 504–535. <https://doi.org/10.1037/0033-2909.134.4.504>.
- Vives-Gabriel, J., J. Schrempf-Stirling, and D. M. Coraiola. 2024. "Dealing With Organizational Legacies of Irresponsibility." *Academy of Management Perspectives* 38, no. 3: 286–303. <https://doi.org/10.5465/amp.2022.0126>.
- Vives-Gabriel, J., W. Van Lent, and F. Wettstein. 2023. "Moral Repair: Toward a Two-Level Conceptualization." *Business Ethics Quarterly* 33, no. 4: 732–762. <https://doi.org/10.1017/beq.2022.6>.
- Vranka, M., and P. Houdek. 2024. "Moral Hypocrisy and the Dichotomy of Hypothetical Versus Real Choices in Prosocial Behavior." *Journal of Economic Psychology* 105: 102772. <https://doi.org/10.1016/j.joep.2024.102772>.
- Walker, M. U. 2006. *Moral Repair: Reconstructing Moral Relations After Wrongdoing*. Cambridge University Press.
- Wallace, H. M., J. J. Exline, and R. F. Baumeister. 2008. "Interpersonal Consequences of Forgiveness: Does Forgiveness Deter or Encourage Repeat Offenses?" *Journal of Experimental Social Psychology* 44, no. 2: 453–460. <https://doi.org/10.1016/j.jesp.2007.02.012>.
- Watkins, Jr., C. Edward, S. H. Reyna, M. J. Ramos, and J. N. Hook. 2015. "The Ruptured Supervisory Alliance and Its Repair: On Supervisor Apology as a Reparative Intervention." *Clinical Supervisor* 34, no. 1: 98–114. <https://doi.org/10.1080/07325223.2015.1015194>.
- Waytz, A., J. Dungan, and L. Young. 2013. "The Whistleblower's Dilemma and the Fairness–Loyalty Tradeoff." *Journal of Experimental Social Psychology* 49, no. 6: 1027–1033. <https://doi.org/10.1016/j.jesp.2013.07.002>.
- Wenzel, M., C. Harous, M. Cibich, and L. Woodyatt. 2023. "Does Victims' Forgiveness Help Offenders to Forgive Themselves? The Role of Meta-Perceptions of Value Consensus." *Journal of Experimental Social Psychology* 105: 104433. <https://doi.org/10.1016/j.jesp.2022.104433>.
- Woodyatt, L., M. Wenzel, T. G. Okimoto, and M. Thai. 2022. "Interpersonal Transgressions and Psychological Loss: Understanding Moral Repair as Dyadic, Reciprocal, and Interactionist." *Current Opinion in Psychology* 44: 7–11. <https://doi.org/10.1016/j.copsyc.2021.08.018>.
- Yang, S., S. Chen, and Y. Liu. 2025. "Perceptions of Economic Fairness Positively Affect Altruistic Punishment." *European Journal of Social Psychology* 55, no. 5: 824–839. <https://doi.org/10.1002/ejsp.3171>.
- Ye, Z., G.-J. Lelieveld, M. K. Noordewier, and E. van Dijk. 2023. "So You Want Me to Believe You're Happy or Angry? How Negotiators Perceive and Respond to Emotion Deception." *Group Decision and Negotiation* 32, no. 6: 1469–1496. <https://doi.org/10.1007/s10726-023-09850-0>.
- Ye, Z., G.-J. Lelieveld, and E. van Dijk. 2025. "Evaluating Negotiators Who Deceptively Communicate Anger or Happiness: On the Importance of Morality, Sociability, and Competence." *Journal of Business Ethics* 199, no. 4: 799–817. <https://doi.org/10.1007/s10551-024-05824-7>.
- Zhang, C., and X. Wei. 2024. "Differentiating the Effects of Power and Status on Unethical Behavior: A Moderated Mediation Meta-Analysis." *Journal of Business and Psychology* 39, no. 4: 871–896. <https://doi.org/10.1007/s10869-023-09919-2>.
- Zhang, L., and A. Ortmann. 2014. "The Effects of the Take-Option in Dictator-Game Experiments: A Comment on Engel's (2011) Meta-Study." *Experimental Economics* 17, no. 3: 414–420. <https://doi.org/10.1007/s10683-013-9375-7>.

Supporting Information

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

Supporting File 1: ejsp70074-sup-0001-SuppMat.docx **Supporting File 2:** ejsp70074-sup-0002-SuppMat.docx