Peer Learning, Enforcement, and Reputation[*]

Yi Chen[†] Kai Du[‡] Phillip Stocken[§] Zhe Wang[¶]

July 13, 2023

Abstract

We consider a two-period collective experimentation model featuring self-interested agents that test a sequentially-rational principal's enforcement propensity through their misconduct, whereby the principal tries to build a reputation for strict enforcement. We find that peer learning between the agents yields more misconduct than when misconduct and enforcement is not observable between the agents, because the principal's reputation concern creates enforcement externalities between them. When the principal's reputation concerns are sufficiently strong, these enforcement externalities dominate the information externalities and heighten misconduct. Whereas reputation always benefits the principal in the opaque setting, it often backfires in the presence of peer learning.

Keywords: Experimentation; Peer Learning; Reputation; Enforcement Externalities; Enforcement Transparency

JEL Classification: D82, D83

[*] We thank Mark Bagnoli, Tim Baldenius, Marco Battaglini, Jeremy Bertomeu, Oliver Boguth, Sam Bonsall, Dongkyu Chang, Jeffrey Coles, Will Cong, Jess Cornaggia, Kimberly Cornaggia, Snehal Banerjee, Jonathan Glover, Matthew Gustafson, Moritz Hiemann, Johannes Hörner, Steve Huddart, Peter Iliev, Yuhta Ishii, Justin Johnson, SP Kothari, Nicholas Lambert, Jorge Lemus, Stefan Lewellen, Fei Li, Lizhi Liu, Qingmin Liu, Christian Lundblad, Nadya Malenko, Andrew McClennan, Brian Mittendorf, Christine Parlour, Heikki Rantakari, Haresh Sapra, Stefan Schantl, Georg Schneider, Timothy Simin, Joel Sobel, Bruno Strulovici, Shyam Sunder, Martin Szydlowski, Juuso Välimäki, Mihail Velikov, Shuyang Wang, Mao Ye, Yao Zeng, Amir Ziv, Alexey Zhdanov, and workshop or conference participants at Penn State University, Columbia Business School, Aalto University, University of North Carolina, 2022 Asian Meeting of Econometric Society, 2022 Conference on Mechanism and Institution Design, 2022 Queen's Organizational Economics Conference for comments and suggestions, and the 13th Accounting Research Workshop at University of Zurich for comments and suggestions. We acknowledge financial support from our respective institutions.

[†] Johnson College of Business, Cornell University; Email: yc2535@cornell.edu.

[‡] Smeal College of Business, Penn State University; Email: kxd30@psu.edu.

[§] Tuck School of Business, Dartmouth College; Email: phillip.c.stocken@tuck.dartmouth.edu.

[¶] Smeal College of Business, Penn State University; Email: zxw192@psu.edu.

1 Introduction

Tension is ubiquitous between self-interested economic agents and the principal who monitors them. Firms may prioritize their own profits at the expense of social welfare, such as in cases of pollution or violations of securities and data protection regulations, while regulators expend resources trying to contain such behavior. Analogously, in a multi-divisional firm setting, divisional managers may implement sub-optimal business plans given their private benefits from empire-building. Although the CEO can intervene and force a division manager to drop the investment plan and avoid the loss, this intervention is typically costly to the CEO. In both cases, as enforcement is costly, some misconduct may go unpunished.

When there are repeated interactions, agents will be interested in learning the principal's enforcement cost—a parameter governing the principal's propensity to enforce. When engaging in misbehavior for their own benefit, these agents simultaneously test the principal's forbearance. Aware of this, the principal aims to build a reputation for aggressive intervention early on to deter future misconduct. Moreover, the sharing of information between the agents may lead to direct and indirect externalities among them. In this light, we ask two questions. Should the principal publicize a case of enforcement and the associated level of misconduct to make an example of the nefarious agent, or keep the case secret? Does a more patient principal deter misconduct more effectively?

To answer these questions, we employ a two-period game featuring a sequentially rational principal and two agents who may misbehave at the principal's expense. In each period, the agents simultaneously choose their levels of misconduct, and observing the misconduct, the principal decides whether to discipline each of them. If the principal tolerates an agent's misconduct, the agent enjoys some private benefit while the principal bears some damage, both proportional to the misconduct level. Alternatively, if the principal disciplines an agent at some fixed cost, then the agent's misconduct is corrected, eliminating any benefit to the agent and damage to the principal. This enforcement cost is the principal's private information; agents only hold a common prior belief at the beginning of the game.

We compare two information structures that vary in terms of how much agents can observe about their peers. One is the opaque benchmark model, where an agent cannot observe a peer's misconduct or the outcome of any enforcement actions taken against the peer. As a result, peer learning is not possible, and the game is reduced to one between a single agent and the principal. In this game, a higher level of misconduct exposes the agent to a greater risk of enforcement but generates more private benefits, when the misconduct is tolerated. Moreover, more egregious misconduct in the first period, if tolerated, emboldens the agent to engage in even more misconduct in the second period, as the agent learns that the principal has a higher enforcement cost. Therefore, an agent's optimal misconduct level in the first period balances the three forces: enforcement risk, private benefit of misconduct, and option value of learning. Concurrently, the *strategic* principal has a reputation concern. She takes into account not only the intra-period costs and benefits of enforcement but also how her current enforcement actions signal her willingness to suppress future misconduct. This reputation concern leads to overzealous enforcement in the first period.

The opaque benchmark model has a unique equilibrium. The principal treats the agents separately and disciplines an agent only when his misconduct is sufficiently high. Importantly, the enforcement criterion does not depend on the peer agent's misconduct. Hence, each agent chooses a strictly positive misconduct level in the first period and then adjusts it upward or downward in the second period given their own enforcement outcome in the first period.

The main model, on the other hand, features an information structure where each agent can also observe their peers' misconduct and enforcement outcomes. While the agents' trade-offs from the benchmark model still apply, agents can also free-ride on the information they learn from their peers, and thus, they curb their own misconduct. This is the canonical effect of *positive information externalities* explored in the collective experimentation literature.

This is not the end of the story, however. Since the strategic principal is aware that her enforcement actions against both agents are publicly observed, she inevitably conditions her enforcement action against one agent on the misconduct of *both* agents. The key intuition for this interdependence is that the principal's cost of mimicking an intolerant type depends on the *proximity* of the agents' misconduct levels in the first period. When the misconduct levels are close, the principal never punishes only the slightly more nefarious agent but not the other, as doing so will allow the agents to infer that the principal has an enforcement cost in a narrow range. The principal is then disabled from pooling with lower-cost types to deter misconduct in the second period. In contrast, when the two agents' misconduct levels are far apart, only disciplining the agent with higher misconduct still enables the principal to "make an example" of the agent, signaling her tough stance without revealing too much information about her enforcement cost.

This interdependence of enforcement decisions creates endogenous *enforcement externalities* between the agents, as one agent's misconduct affects the other's chance of being disciplined. The enforcement externalities can be positive or negative. They are *positive* if they *discourage* an agent from engaging in misconduct as he shields behind his peer's misconduct. Alternatively, they are *negative* if they *encourage* an agent to increase his misconduct because part of his heightened enforcement risk is shared with his peer.

The main model, where peer misconduct and enforcement are transparent, has a unique equilibrium. It takes starkly different forms depending on the principal's patience. A relatively *impatient* principal is less concerned about her future reputation resulting from the enforcement decisions in the first period, in particular, the unfavorable inference from punishing only one agent. Therefore, the principal will sometimes discipline only one agent if his misconduct is much higher than the other's. This creates an incentive for each agent to reduce their misconduct to shield behind their peer. In this case, the enforcement externalities, which are positive, work in the *same* direction as the positive information externalities, further discouraging misconduct and benefiting the principal.

In contrast, a relatively *patient* principal is so concerned about her future reputation that she wishes to avoid the unfavorable inference from punishing only one agent. In essence, her hands are tied as she is compelled to treat the two agents similarly and condition her collective enforcement decision on the *total* misconduct levels. Thus, although being more patient encourages the principal to discipline misconduct more aggressively, this “both or neither” enforcement strategy backfires as the risk of enforcement is shared between the agents, thereby encouraging misconduct. These enforcement externalities, which are negative, work in the opposite direction to the positive information externalities and dominate them. The outcome is thus heightened misconduct, which harms the principal.

Our analysis offers two key sets of implications. First, we highlight the effect of peer learning on equilibrium misconduct. When the principal is relatively impatient, peer learning discourages misconduct and thereby reduces the principal’s cost. This prediction echoes the traditional literature on collective experimentation (Bolton and Harris, 1999; Keller et al., 2005), which emphasizes information spillovers between the agents. When the principal is relatively patient, however, peer learning *encourages* misconduct, which harms the principal. This is because a more patient principal becomes overly concerned about her reputation and is hesitant to punish only one agent. This constrains the principal to treat the agents as a group, causing negative enforcement externalities to arise between the agents and inducing heightened misconduct. These novel endogenous enforcement externalities, uncovered in our model, are strong enough to dominate the information spillovers, thereby reversing the prediction in the extant collective experimentation literature.

Second, we illustrate how the principal’s reputation concern affects misconduct. When information is opaque between agents (i.e., in a single-agent scenario), a more patient principal always better deters misconduct, which is largely consistent with the conventional wisdom about reputation. However, when learning is possible between agents, the misconduct levels are non-monotonic in the principal’s patience. On the one hand, a higher level of patience gives the principal stronger incentives to build a tough reputation, resulting in more stringent enforcement that quashes misconduct. On the other hand, as the principal becomes more patient, she becomes more concerned about her reputation and hesitates to discipline just one agent. Instead, she is trapped into treating both agents as a group, resulting in higher misconduct levels that harm the principal.

Finally, in extensions of our model, we illustrate the robustness of our main results with (a) alternative parameter ranges, (b) heterogeneous agents, and (c) more than two periods.

Related Literature Our paper belongs to the literature on experimentation (Rothschild, 1974; Aghion et al., 1991; Manso, 2011; Bergemann and Hege, 2005; Nanda and Rhodes-Kropf, 2017). See Bergemann and Välimäki (2006) and Hörner and Skrzypacz (2016) for surveys of this literature. Specifically, a branch of literature studies collective experimentation where multiple agents learn from each other. The literature typically assumes that the object about which the players learn, or the "bandit arm," is *non-strategic*, and consequently the information externalities among players drive the dynamics of learning (e.g., Bolton and Harris, 1999; Keller et al., 2005; Bonatti and Hörner, 2011; Chen, 2020). In our paper, in contrast, agents learn about a *strategic* principal who optimally manages her reputation. We find that peer learning among agents not only results in free-riding of information, but also creates endogenous enforcement externalities. For a sufficiently patient principal, we show that this novel effect dominates the free-riding effect and raises equilibrium experimentation through misconduct. Bergemann and Välimäki (2000) is among the few papers that study buyers' collective learning when facing strategic sellers. This learning is about the unknown quality of a new product in a setting where information asymmetry is absent. In contrast, our paper considers a one-sided learning model, where agents experiment to elicit the principal's private information whereas the principal tries to jam the signal. Marinovic and Szydlowski (2022) investigate an environment where one or more agents and a monitor learn about the monitor's unknown ability. They find that when the monitor can detect misconduct, but not the identity of the misbehaving agent, the monitor has to collectively punish all agents. The agents can free-ride on this collective punishment, which leads to heightened misconduct. In our work, in contrast, the principal observes individual misconduct and can choose individual punishments, although sometimes she will *endogenously* choose an enforcement strategy that resembles collective punishment. When she does so, misconduct is either encouraged or discouraged, depending on the patience of the principal.

Our paper is also related to the literature on the counter-productive effect of reputation, i.e., bad reputation. Ely and Välimäki (2003) uncover the bad reputation effect in repeated games where a sufficiently patient long-run player becomes so obsessed with its reputation that it fails to take the correct stage action and thus is ousted by the short-run opponents. Corona and Randhawa (2010) consider a setting with a manager and a monitor who can detect the manager's fraud. They find that the reputation concerns of the monitor may heighten the monitor's reluctance to uncover fraud for fear of revealing previously undetected fraud. Deb, Mitchell, and Pai (2022) study an experimentation setting with an informed agent who cares about its reputation for being a good bandit arm that benefits the principal. They show that the agent's reputational incentives are so strong that they can destroy the relationship between the principal and the agent, unless they can coordinate on inefficient strategies. Unlike these papers that feature bad reputation with only one agent (or when agents cannot learn from each other), our model displays a bad reputation effect when multiple agents learn from each other. Bar-Isaac and Deb (2014) study a repeated reputation game between one agent and two audiences with *diverse* preferences, where the agent's persistent private information is its preference alignment with one of the audiences. They show that transparency between the audiences improves welfare. In contrast, our model features *homogeneous* audiences, and we find that transparency can be harmful.

Our results on the effect of information transparency and the principal's patience shed light on questions in various institutional settings. Our paper is related to the literature on regulatory reputation management (Boot and Thakor, 1993; Morrison and White, 2013; Shapiro and Skeie, 2015; Huang, 2017), regulatory enforcement transparency (see Goldstein and Sapra (2013) and Goldstein and Leitner (2020) for surveys of the literature that focuses on banking regulations), and crime (Bond and Hagerty, 2010). We contribute to this literature by exploring a novel channel through which the reputation-building motive of a strategic principal interacts with peer-learning among agents to yield guidance on whether a principal should publicize the agents' misconduct and enforcement.

The paper proceeds as follows. Section 2 introduces the model. Section 3 analyzes a benchmark environment where peer misconduct and enforcement are opaque. Section 4 characterizes the equilibrium and derives implications when peer misconduct and enforcement are transparent. Section 5 extends the analysis. Section 6 concludes. All proofs are relegated to the Appendix.

2 Model

Consider a model of collective experimentation between two agents (with the pronoun he) facing a strategic principal (with the pronoun she). The game has two periods. In each period $t = 1, 2$, agents $i = A, B$ simultaneously choose the levels of *misconduct*, $x_t^i \geqslant 0$. After observing (x_t^A, x_t^B), the principal makes *enforcement* decisions $(e_t^A, e_t^B) \in \{0, 1\}^2$, such that $e_t^i = 1$ denotes the decision to discipline agent i in period t and $e_t^i = 0$ indicates otherwise.

If agent i's misconduct x_t^i is tolerated ($e_t^i = 0$), the agent derives private benefits normalized to x_t^i while the principal internalizes proportional damages kx_t^i, where the coefficient $k > 0$ is interpreted as the principal's damage sensitivity. Alternatively, if the agent is disciplined ($e_t^i = 1$), the agent's private benefit is eliminated and the principal avoids the damage. Additionally, the agent suffers a *penalty* $L > 0$,[1] and the principal incurs an *enforcement cost* $c \geqslant 0$. The enforcement cost c and damage sensitivity k are the principal's private information and persist across periods. At the beginning of period 1, agents hold a common prior about the joint distribution of (c, k).

We compare two information structures. In the opaque benchmark (Section 3), one

[1] While we assume L to be an exogenous number in this paper, the qualitative results still hold when $L(\cdot)$ is a function of own misconduct.

agent's misconduct and enforcement outcome are not observed by the other agent, whereas in the main model (Section 4), the misconduct levels (x_1^A, x_1^B) and enforcement outcomes (e_1^A, e_1^B) in period 1 are common knowledge between the agents at the end of period 1. After observing the available information, the agents enter period 2 with Bayes-updated beliefs. The timeline is summarized in Figure 1.

All players are risk-neutral. The principal has discount factor $\delta_P \in (0,1)$, whereas the agents have a common discount factor $\delta \in (0,1)$. A higher discount factor is interpreted as the player being more patient. Given the realized misconduct levels x_t^i and enforcement decisions e_t^i, $t = 1, 2$, agent i enjoys a *total discounted payoff* across both periods of:

$$U^i \equiv \big[x_1^i(1-e_1^i) - Le_1^i\big] + \delta\big[x_2^i(1-e_2^i) - Le_2^i\big]. \tag{1}$$

The principal incurs a *total discounted cost* across both periods of:

$$\begin{aligned} C &\equiv \sum_{i=A,B} \Big(\big[kx_1^i(1-e_1^i) + ce_1^i\big] + \delta_P\big[kx_2^i(1-e_2^i) + ce_2^i\big]\Big) \\ &= k\sum_{i=A,B} \Big(\big[x_1^i(1-e_1^i) + \theta e_1^i\big] + \delta_P\big[x_2^i(1-e_2^i) + \theta e_2^i\big]\Big), \end{aligned}$$

where:

$$\theta \equiv \frac{c}{k} \geqslant 0$$

is interpreted as the principal's *normalized enforcement cost*. In our linear and additive framework, the model is strategically equivalent to a normalized one with $k = 1$ and $c = \theta$, where θ summarizes the principal's private information. Hereafter, we let $k = 1$ for notational convenience, and refer to θ as the principal's *type*. A higher θ corresponds to a lower enforcement propensity and hence a more tolerant principal. Agents have common prior beliefs at the beginning of period 1 that θ is continuously distributed with c.d.f. F and p.d.f. f on the support $[0, \infty)$. We require the parameters to satisfy the following conditions.

Assumption 1 (Parameters)

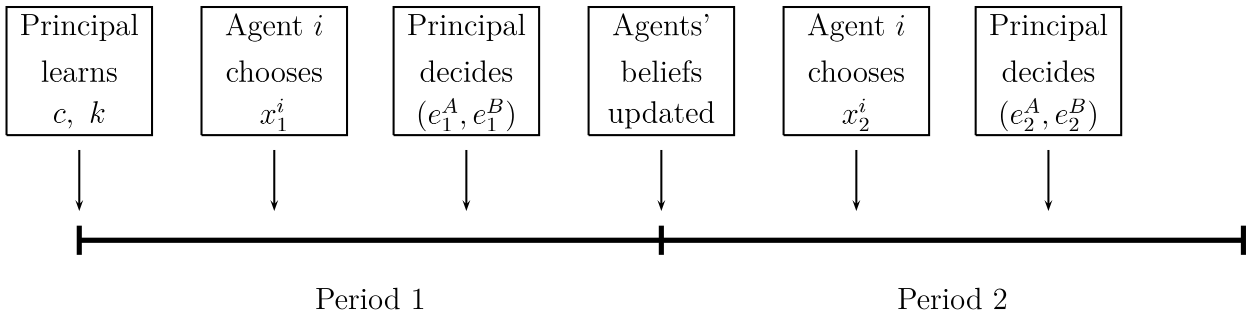

Principal learns $c,\ k$
Agent i chooses x_1^i
Principal decides (e_1^A, e_1^B)
Agents' beliefs updated
Agent i chooses x_2^i
Principal decides (e_2^A, e_2^B)
Period 1
Period 2

Figure 1: Timeline of the two-period game

(i) *The hazard rate* $h(\theta) \equiv \frac{f(\theta)}{1-F(\theta)}$ *is continuously increasing in* θ*.*

(ii) *The penalty is moderate:* $1 \leqslant Lh(0) < 1 + \delta - \delta_P$*.*

Part (i) is a mildly restrictive monotone hazard rate property. Part (ii) requires the penalty to be large enough, i.e., $L \geqslant 1/h(0)$, so that if there was only one period, the agents would try to avoid any enforcement risk.[2] On the other hand, the penalty is sufficiently low, i.e., $L < (1+\delta-\delta_P)/h(0)$, such that the agents will engage in some misconduct in period 1. Note that (ii) implies $\delta_P < \delta$, that is, the principal is more impatient than the agents. Otherwise, the principal always punishes any agent with positive misconduct in period 1 to signal her type, and foreseeing this, agents are deterred from any misconduct in period 1.

With slight abuse of notation, the strategy of agent i is a pair (x_1^i, x_2^i), where $x_1^i \in [0, \infty)$. In the opaque benchmark, x_2^i maps (x_1^i, e_1^i) into $[0, \infty)$, while in the main model, x_2^i maps $\{x_1^j, e_1^j\}_{j=A,B}$ into $[0, \infty)$. The principal's strategy is a pair (e_1, e_2), where e_1 maps $\big(\theta, \{x_1^j\}_{j=A,B}\big)$ into $\{0,1\} \times \{0,1\}$ and e_2 maps $\big(\theta, \{x_1^j, e_1^j, x_2^j\}_{j=A,B}\big)$ into $\{0,1\} \times \{0,1\}$. We impose the following constraint on the principal's strategy:

Assumption 2 (Enforcement Strategy)

If the principal is indifferent between disciplining an agent or not, she breaks the tie by not disciplining. Formally, for any principal type θ*, period* $t = 1, 2$ *and agent* $i = A, B$*, the principal's enforcement decision* e_t^i *is lower semi-continuous in* (x_t^A, x_t^B)*.*

[2] We show numerically in Section 5.2 that the key driving forces of the model remain unchanged if this assumption is relaxed.

This assumption ensures that agents' best responses are always well-defined. It is without loss of generality because an agent can always reduce his misconduct by ε to avoid enforcement.

We characterize the perfect Bayesian equilibrium (PBE). A PBE is a strategy profile $\{x_t^A, x_t^B, e_t\}_{t=1,2}$ along with agent i's ($i = A, B$) posterior belief $q^i \in \Delta([0, \infty))$ at the beginning of period 2, such that: (a) agent i's strategy (x_1^i, x_2^i) maximizes his expected total discounted payoff, denoted $\mathbb{E}(U^i)$, given the strategies of agent $-i$ and the principal; (b) the principal's strategy $\{e_t\}_{t=1,2}$ minimizes her expected total discounted cost, denoted $\mathbb{E}(C)$, given the agents' strategies; and (c) the agent i's posterior belief q^i is consistent with the strategies and Bayes' rule whenever possible.

3 Benchmark: No Peer Learning

We first consider an opaque benchmark in which an agent cannot observe the peer's misconduct nor the enforcement outcome. This opaque information structure eliminates peer learning between the agents. As each agent only relies on his own experience to learn the principal's type and the principal's cost is separable between the agents, the game is equivalent to two replications of a *one-agent-one-principal* game.

Using backward induction, we begin with the continuation game in period 2. The following lemma characterizes the equilibrium strategies of the three parties.

Lemma 1 (Period-2 Equilibrium)

In period 2, for $i = A, B$:

(i) *The principal disciplines agent i if and only if $x_2^i > \theta$.*

(ii) *If the posterior belief q^i has the truncated support $[\theta_L, \theta_H) \subset [0, \infty)$ from the prior, then $x_2^i = \theta_L$.*

Lemma 1 states that all players behave myopically in the terminal period. The principal compares the misconduct damage x_2^i to the enforcement cost θ, without considering her reputation. Analogously, an agent maximizes his within-period payoff as there is no further option value from learning. Given Assumption 1, an agent chooses a misconduct level equal to the lower bound of his posterior belief support, thereby avoiding any enforcement risk; that is, the agent chooses the highest misconduct while still avoiding punishment.

We now turn to period 1. Suppose agent $i \in \{A, B\}$ has chosen misconduct x_1^i. If a type-θ principal disciplines this agent, her discounted cost from agent i alone is:

$$\theta + \delta_P \min\{\theta, x_2^i(x_1^i, e_1^i = 1)\}, \tag{2}$$

which is the sum of the enforcement cost in period 1 and the discounted continuation cost in period 2 when $e_1^i = 1$. If, instead, she does not discipline agent i, the cost is:

$$x_1^i + \delta_P \min\{\theta, x_2^i(x_1^i, e_1^i = 0)\}, \tag{3}$$

which is the damage suffered in period 1 plus the discounted continuation cost in period 2 when $e_1^i = 0$. The principal disciplines agent i in period 1 if and only if (2) is less than (3).

Principal's cost from agent i displays the single-crossing property in θ. That is, given any misconduct x_1^i in period 1, the higher the enforcement cost θ, the less willing the principal is to discipline the agent.[3] Consequently, the principal follows a *cutoff* strategy such that an agent is disciplined if and only if θ is strictly below some cutoff $\theta^\dagger$.

We now determine the equilibrium cutoff type $\theta^\dagger$ of the principal that is indifferent between disciplining the agent and not. When the type-$\theta^\dagger$ principal disciplines the agent, (2) reduces to $\theta^\dagger + \delta_P \cdot 0$. This follows because, according to Lemma 1(ii), the agent chooses $x_2^i = 0$ to avoid any enforcement risk when the principal's type is inferred to be in $[0, \theta^\dagger)$. When the agent is not disciplined, (3) simplifies to $x_1^i + \delta_P \cdot \theta^\dagger$, because with the principal's type inferred to be in $[\theta^\dagger, \infty)$, the agent chooses $x_2^i = \theta^\dagger$. The cutoff $\theta^\dagger$ can then be solved

[3] To see this, observe that the derivative of (2) with respect to θ is no lower than 1, while the derivative of (3) with respect to θ is no higher than $\delta_P < 1$. Thus, the difference between (2) and (3) increases in θ, and the single-crossing property holds. This argument is true for any $x_2^i(\cdot, \cdot)$.

from the indifference between (2) and (3), which depends on x_1^i:

$$\theta^\dagger = \phi(x_1^i) \equiv \frac{x_1^i}{1-\delta_P}. \tag{4}$$

The principal disciplines agent i in period 1 if and only if $\theta < \theta^\dagger$, or equivalently,

$$x_1^i > \theta(1-\delta_P). \tag{5}$$

Understanding the principal's strategy, agent i anticipates the risk associated with each level of period-1 misconduct and optimally chooses x_1^i to maximize the expected payoff:

$$\left(x_1^i + \delta \cdot \phi\left(x_1^i\right)\right)\left(1 - F\left(\phi\left(x_1^i\right)\right)\right) + (-L) \cdot F\left(\phi\left(x_1^i\right)\right). \tag{6}$$

To interpret, the principal's type is higher than the cutoff $\theta^\dagger = \phi(x_1^i)$ with probability $1 - F\left(\phi\left(x_1^i\right)\right)$, in which case the agent's period-1 misconduct is tolerated. The agent then enjoys a private benefit of x_1^i in period 1, infers the principal's type to be at least $\phi(x_1^i)$, and chooses period-2 misconduct $x_2^i = \phi(x_1^i)$. With the complementary probability, the principal's type is lower than the cutoff, in which case the agent is disciplined in period 1. The agent then suffers the penalty L in period 1, and chooses $x_2^i = 0$ in period 2 with his updated belief. The first-order condition with respect to x_1^i yields the equilibrium period-1 misconduct, characterized in the next proposition.

Proposition 1 (Equilibrium: No Peer Learning)
Let $\phi(\cdot)$ be defined as in (4). In the absence of peer learning, there is a unique perfect Bayesian equilibrium. Type-θ principal disciplines an agent with x_1^i in period 1 if and only if $\theta < \phi(x_1^i)$ (or equivalently, $x_1^i > \theta(1-\delta_P)$). Agent i's period-1 misconduct is $x_1^i = x^*$, where $x^* > 0$ uniquely solves:

$$\left(L + (1+\delta-\delta_P)\phi(x^*)\right) h\left(\phi(x^*)\right) = 1 + \delta - \delta_P. \tag{7}$$

In addition, agent i's period-2 misconduct is $x_2^i = 0$ if $e_1^i = 1$ and $x_2^i = \phi(x^*)$ if $e_1^i = 0$.

Each agent chooses a strictly positive and identical misconduct x^* in the first period. The second-period misconduct increases to $\phi(x_1^i)$ if the principal tolerates it and drops to zero otherwise. The trade-off an agent faces in period 1 is as follows. With a higher misconduct x_1^i, the agent faces increased risk of enforcement while enjoying greater current benefits conditional on the principal's forbearance. In addition to these intra-period trade-offs, higher x_1^i also enhances the option value of learning: if the higher misconduct is tolerated in period 1, the principal is revealed to be even weaker than the agent initially supposed, causing the agent to raise his period-2 misconduct. As (4) reflects, a higher x_1^i maps to a higher cutoff type $\phi(x_1^i)$, which is exactly the period-2 misconduct. Thus, an agent has an extra *experimentation* incentive to probe the principal's enforcement propensity.

The principal, meanwhile, anticipates the agents' experimentation and thus disciplines them aggressively in period 1 to build a tough reputation. This tendency is reflected in (5) as a type-θ principal disciplines an agent with misconduct as low as $\theta(1 - \delta_P)$, which is below the misconduct of θ that the principal tolerates when she is myopic. Moreover, the principal's period-1 enforcement strategy is *independent* across agents, that is, the decision whether to discipline one agent depends only on that agent's misconduct and not on his peer's.

We also investigate whether a principal with higher patience is better at deterring misconduct. Intuitively, as the principal becomes more patient, she cares more about her reputation because it determines her period-2 cost. Accordingly, she adopts a more stringent enforcement criterion in the first period to build a tough reputation. Anticipating this aggressiveness, the agents curb their misconduct. The next corollary formalizes this intuition.

Corollary 1 (Reputation Effect without Peer Learning)

A higher δ_P strictly reduces the agents' misconduct in both periods.

4 Main Model: Peer Learning

We now turn to the main setting in which an agent observes not only the peer's misconduct but any enforcement against the peer. For an agent, the trade-offs identified in the opaque benchmark still apply. Unlike the benchmark, however, the peer-learning now introduces two forms of *externalities.* First, an agent can free-ride on the information revealed by his peer's experience, thereby lowering his own misconduct. This leads to positive *information externalities.* Second, the principal's enforcement against one agent may depend on the other agent's misconduct, which endogenously generates *enforcement externalities* between them.

Using backward induction, we start with period 2. The characteristics of any equilibrium identified in Lemma 1 continue to apply to the main model because all three players act myopically in period 2.

4.1 Period 1: Principal's Enforcement

Next, we consider the first period. Fixing an arbitrary misconduct profile (x_1^A, x_1^B), we characterize properties of the principal's strategy that necessarily arise in equilibrium.

Lemma 2 (Properties of Enforcement)
(i) (Monotonicity) *In period 1, if* $x_1^i > x_1^j$ *and* $e_1^j = 1$*, then* $e_1^i = 1$*.*
(ii) (Cutoff Strategy) *Given any misconduct profile* (x_1^A, x_1^B)*, there exist cutoff types* $0 \leqslant \theta^{**} \leqslant \theta^* < \infty$ *such that the principal disciplines both agents when* $\theta < \theta^{**}$*, disciplines one agent when* $\theta^{**} \leqslant \theta < \theta^*$*, and disciplines neither agent when* $\theta \geqslant \theta^*$*.*

Lemma 2(i) implies that if the principal disciplines only one agent, then it must be the one with the higher level of misconduct. Part (ii) shows that the principal follows a *cutoff* strategy that depends on her type θ, echoing the counterpart in the benchmark.[4] As the principal's type θ increases, her enforcement strategy becomes increasingly tolerant.

[4] As in the benchmark, the fact that the principal follows a cutoff enforcement strategy does not depend on the agents' second-period strategy.

Although the principal's cutoff strategy contains three cases, it is possible that $\theta^{**} = \theta^*$ for some (x_1^A, x_1^B). In this degenerate scenario, the principal disciplines both agents if her type is low and disciplines neither otherwise, but never finds it optimal to discipline only one agent. This scenario where the principal punishes both or neither arises when the agents' misconduct x_1^A and x_1^B are *close* together.

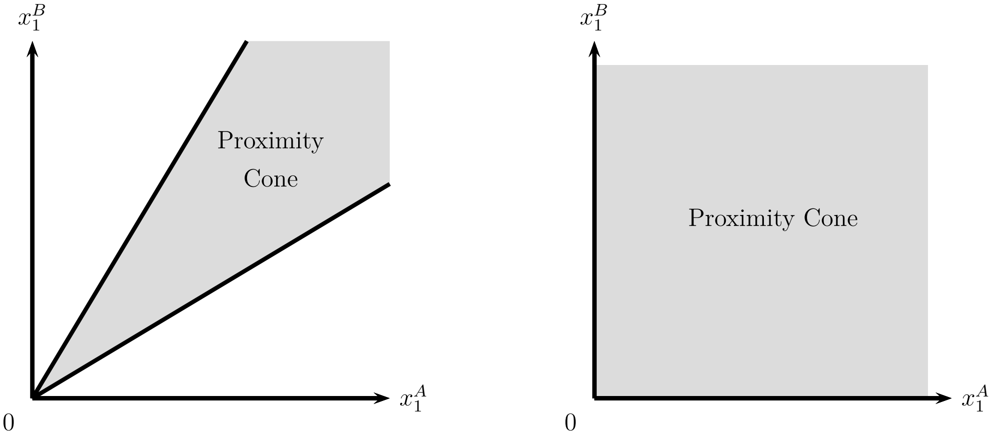
The proximity of the agents' misconduct levels is vital for determining how the principal develops a reputation for strict enforcement. The principal's reputation building, and hence enforcement strategy, depends on whether the agents' misconduct levels are *close* or *far apart.* On the one hand, suppose the two agents' period-1 misconduct levels are close, where x_1^A is only *slightly* higher than x_1^B. If the principal punishes only agent A but not B, the reputation building is limited as the principal will be considered as having an enforcement cost not low enough to justify punishing both agents, and worse still, her enforcement cost is revealed to lie in a narrow range, leaving her little room to pool with lower-cost types. On the other hand, suppose agent A's misconduct is *far above* that of agent B. This time, punishing only agent A does not substantially hurt the principal's reputation. The large gap in the agents' misconduct levels enables the principal to pool with much lower cost types, despite having punished only one agent.

Formally, we define the *proximity cone* as a subset of the x_1^A–x_1^B plane, where the principal never disciplines only one agent when the actions (x_1^A, x_1^B) fall inside this cone.

Definition 1 (Proximity Cone)
For $\delta_P \in (0,1)$, the proximity cone *is the set:*

$$\mathcal{P}(\delta_P) \equiv \left\{(x_1^A, x_1^B) \in [0,\infty)^2 : \; x_1^B \geqslant x_1^A(1-2\delta_P), \; x_1^A \geqslant x_1^B(1-2\delta_P)\right\}.$$

Figure 2(a) illustrates the proximity cone for $\delta_P < \frac{1}{2}$, which is the area sandwiched between the two rays $x_1^B = x_1^A(1-2\delta_P)$ and $x_1^B = \frac{x_1^A}{1-2\delta_P}$. These rays correspond to the misconduct profiles for which punishing only one agent becomes a dominated strategy for



(a) $\delta_P < \frac{1}{2}$

(b) $\delta_P \geqslant \frac{1}{2}$

Figure 2: Proximity cone for small and large δ_P.

all principal types. As δ_P grows, the cone expands and, when $\delta_P \geqslant \frac{1}{2}$, covers the entire quadrant, as Figure 2(b) shows.

The next proposition describes the principal's enforcement strategy for misconduct profiles inside and outside the proximity cone.

Proposition 2 (Principal Strategy in Period 1)

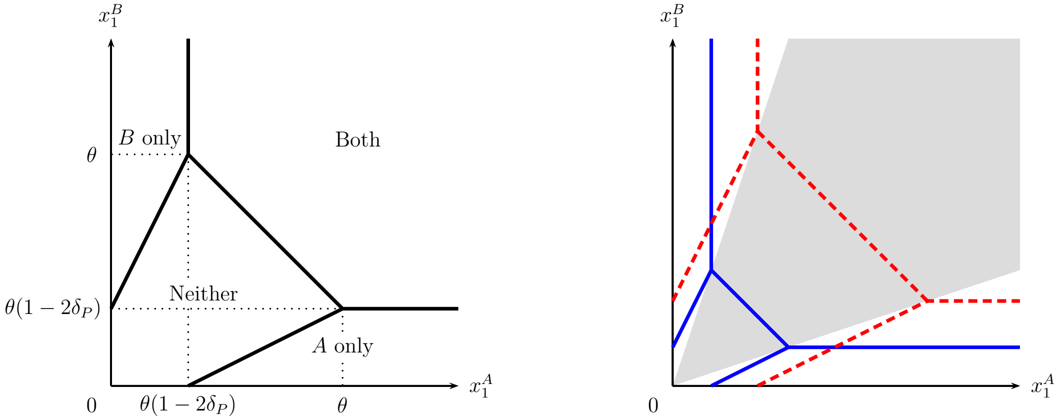
Define:

$$\begin{array}{rcll} \phi_{1|2}(x_1^A, x_1^B) & = & \dfrac{\min\{x_1^A, x_1^B\}}{1-2\delta_P} & \text{for } \delta_P \in \left[0, \tfrac{1}{2}\right), \\ \phi_{0|1}(x_1^A, x_1^B) & = & \dfrac{\max\{x_1^A, x_1^B\}(1-2\delta_P) - \min\{x_1^A, x_1^B\}(2\delta_P)}{(1-2\delta_P)^2} & \text{for } \delta_P \in \left[0, \tfrac{1}{2}\right), \\ \phi_{0|2}(x_1^A, x_1^B) & = & \dfrac{x_1^A + x_1^B}{2(1-\delta_P)} & \text{for } \delta_P \in [0, 1) . \end{array} \tag{8}$$

(i) *If (x_1^A, x_1^B) is in the proximity cone, then the principal disciplines both agents when $\theta < \phi_{0|2}(x_1^A, x_1^B)$ and disciplines neither otherwise.*

(ii) *If (x_1^A, x_1^B) is not in the proximity cone, then the principal disciplines both agents when $\theta < \phi_{1|2}(x_1^A, x_1^B)$, disciplines one agent when $\phi_{1|2}(x_1^A, x_1^B) \leqslant \theta < \phi_{0|1}(x_1^A, x_1^B)$, and disciplines neither when $\theta \geqslant \phi_{0|1}(x_1^A, x_1^B)$.*

We describe the enforcement strategy graphically by characterizing the enforcement



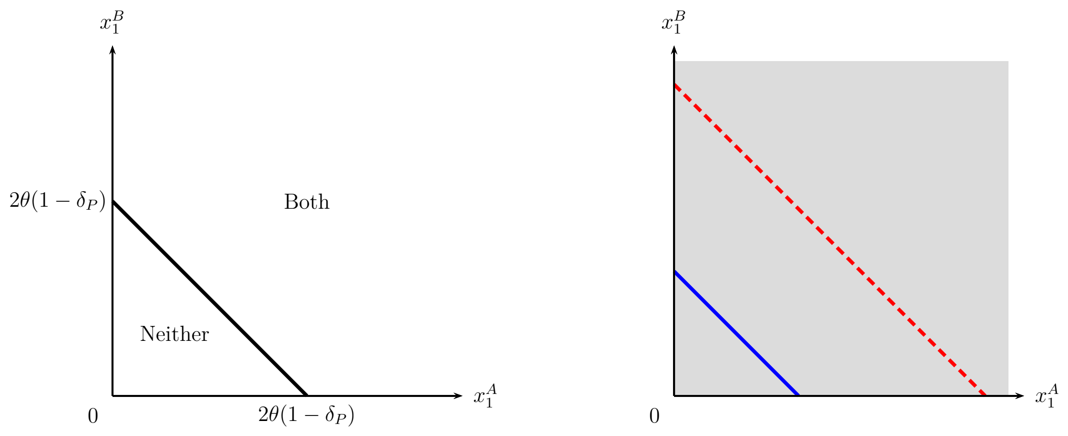
(a) Enforcement zones: fixed θ

(b) Enforcement zones: varying θ

Figure 3: Principal's enforcement zones in period 1 for $\delta_P < \frac{1}{2}$. (a): four distinct enforcement zones for a fixed θ. (b): solid (resp. dashed) lines are the boundaries of the enforcement zones for a low (resp. high) type, and the shaded area is the proximity cone.

zones on the x_1^A–x_1^B plane. These enforcement zones take different shapes, depending on whether the principal is relatively impatient or patient. For an *impatient* principal ($\delta_P < \frac{1}{2}$), Figure 3(a) illustrates the four distinct enforcement zones when the principal's type θ is fixed. The solid lines are the boundaries between adjoining zones; the principal is indifferent between adjacent decisions at these boundaries. Although it is still true that a type-θ principal disciplines an agent whenever his misconduct level is too high, the key difference from the opaque benchmark is that the criterion for "too high" now depends on the misconduct of the other agent. As the principal's type θ increases in Figure 3(b), the enforcement zones scale up from a lower type (solid lines) to a higher type (dashed lines). The higher type is more lenient because, for any profile (x_1^A, x_1^B), she disciplines weakly fewer agents than a lower type. Notably, by varying θ, the boundary between punishing neither and both agents, which has a slope of -1, traces out exactly the proximity cone (shaded area).

We now turn to a *patient* principal, defined by $\delta_P \geqslant \frac{1}{2}$. In this case, the proximity cone covers the entire quadrant, as the principal always bundles the two agents together for enforcement. Accordingly, fixing the principal's type, there are only two enforcement



(a) Enforcement zones: fixed θ

(b) Enforcement zones: varying θ

Figure 4: Principal's enforcement zones in period 1 for $\delta_P \geqslant \frac{1}{2}$. (a): two distinct enforcement zones for a fixed θ. (b): the solid (resp. dashed) line is the boundary of the enforcement zones for a low (resp. high) type, and the shaded area is the proximity cone.

zones instead of four, as Figure 4(a) illustrates. As the principal's type changes, Figure 4(b) compares the enforcement zones for a lower and a higher type of principal, where again, the principal becomes more lenient as her type increases.

Collectively, Figure 3 and Figure 4 show that principal's enforcement actions are guided by how closely the agents' misconduct levels align (that is, if they fall within the proximity cone). As the principal becomes more patient, the proximity cone expands. Intuitively, because a more patient principal places a higher value on her reputation, she is more concerned about the potential reputational damage that might arise if she penalizes only one agent. When she is sufficiently patient ($\delta_P \geqslant \frac{1}{2}$), her focus on her reputation in period 2 is so strong that she constrains herself to always punishing both agents in period 1. This can be observed in Figure 4(b), where the proximity cone (illustrated as the shaded area) expands to cover the entire quadrant.

Unlike in the opaque benchmark, the principal's enforcement strategy in the main model is interdependent between the agents. In other words, her equilibrium strategy creates endogenous *enforcement externalities* between the agents. These externalities can be either positive or negative, which we define next.

Definition 2 (Enforcement Externalities) *For $i = A, B$:*

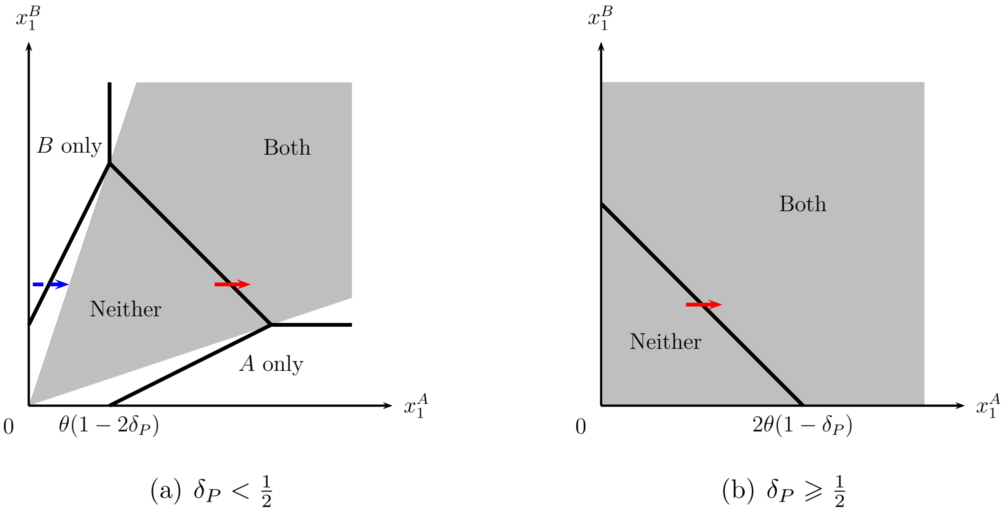
(i) ***Positive*** *enforcement externalities arise when agent i's higher period-1 misconduct benefits peer agent $-i$ by reducing the latter's probability of enforcement in period 1;*

(ii) ***Negative*** *enforcement externalities arise when agent i's higher period-1 misconduct harms peer agent $-i$ by raising the latter's probability of enforcement in period 1.*

If an agent engages in misconduct that is significantly less egregious than that of his peer, he is likely to avoid punishment while exposing his peer. However, if he increases his level of misconduct to a point where it is similar to his peer, he reduces the peer's enforcement risk, creating *positive* enforcement externalities for his peer. This is because the principal now becomes more hesitant to punish the peer agent, as doing so would require punishing both agents, resulting in double the cost. When $\delta_P < \frac{1}{2}$, the dashed arrow in Figure 5(a) illustrates these positive externalities, for a given principal type θ. As agent A's misconduct x_1^A increases and crosses the upward-sloping boundary, the principal switches from disciplining only agent B to disciplining neither agent, reducing agent B's enforcement risk. In other words, agent A's increased misconduct protects agent B from being punished. Notably, these positive externalities only occur outside the proximity cone.

Alternatively, if an agent's misconduct is close to his peer's misconduct, then further raising his misconduct imposes *negative* enforcement externalities on the peer. This is because when the misconduct levels are sufficiently close that they fall in the proximity cone, the principal punishes both agents or neither agent, depending on the *aggregate* misconduct of both agents. Therefore, an agent's higher misconduct raises the sum and exposes his peer to greater risk. For a type-θ principal, the negative externalities can be seen in both panels of Figure 5 as the solid arrows. As x_1^A increases past the boundary between "Neither" and "Both", the principal switches from disciplining neither agent to disciplining both. These negative externalities only arise inside the proximity cone.

The enforcement externalities are novel insights generated from the interaction of the principal's reputation concern and the agents' peer learning. The enforcement externalities



(a) $\delta_P < \frac{1}{2}$

(b) $\delta_P \geqslant \frac{1}{2}$

Figure 5: Enforcement externalities given the principal's type θ. Solid (resp. dashed) arrow shows the negative (resp. positive) externalities. (a): both externalities exist for $\delta_P < \frac{1}{2}$. (b): only negative externality exists for $\delta_P \geqslant \frac{1}{2}$. The shaded area is the proximity cone.

do not exist if either the principal is myopic ($\delta_P = 0$), or in the opaque benchmark where there is no peer learning about the agent. Consequently, without either ingredient, an agent's misconduct has no bearing on its peer's enforcement.

4.2 Period 1: Agents' Misconduct

We now consider the agents' incentives to engage in misconduct given the principal's enforcement strategy. Suppose the agents' misconduct falls *inside* the proximity cone. Agent i's expected discounted payoff is

$$\mathbb{E}(U^i) = \left(x_1^i + \delta \cdot \phi_{0|2}\right)\left(1 - F(\phi_{0|2})\right) + (-L)F(\phi_{0|2}), \tag{9}$$

where U^i is defined in (1), and $\phi_{0|2}$ is short for $\phi_{0|2}(x_1^A, x_1^B)$ defined in (8). To interpret, when facing misconduct profile (x_1^A, x_1^B) inside the proximity cone, $\phi_{0|2}$ is the cutoff type between disciplining both agents and disciplining neither. When the principal disciplines neither agent in period 1 (which occurs with probability $1 - F(\phi_{0|2})$), agent i enjoys private benefit x_1^i and infers the principal's type to be at least $\phi_{0|2}$. Applying Lemma 1, the agent chooses $x_2^i = \phi_{0|2}$ in period 2. When, however, the principal disciplines both agents in period 1, agent i suffers the penalty L and refrains from any misconduct in period 2. Expression (8) shows that $\phi_{0|2}$ is proportional to the aggregate misconduct $x_1^A + x_1^B$. A higher x_1^i increases this aggregate misconduct, thereby raising the enforcement risk for *both* agents. In other words, the enforcement risk for each agent, represented by $F(\phi_{0|2}) = F(\frac{x_1^i + x_1^{-i}}{2(1-\delta_P)})$, increases in x_1^i. Thus, within the proximity cone, the agents' actions create *negative* enforcement externalities.[5]

Alternatively, suppose the agents' misconduct falls *outside* of the proximity cone, that is, one agent's misconduct is much higher than his peer's. If agent i is *leading* in misconduct, that is, $x_1^i > x_1^{-i}$, then his expected discounted payoff is

$$\mathbb{E}(U^i) = \left(x_1^i + \delta \cdot \phi_{0|1}\right)\left(1 - F(\phi_{0|1})\right) + \left(-L + \delta \cdot \phi_{1|2}\right)\left(F(\phi_{0|1}) - F(\phi_{1|2})\right) + (-L)F(\phi_{1|2}). \quad (10)$$

If, instead, agent i is *trailing*, that is, $x_1^i < x_1^{-i}$, then

$$\mathbb{E}(U^i) = \left(x_1^i + \delta \cdot \phi_{0|1}\right)\left(1 - F(\phi_{0|1})\right) + \left(x_1^i + \delta \cdot \phi_{1|2}\right)\left(F(\phi_{0|1}) - F(\phi_{1|2})\right) + (-L)F(\phi_{1|2}). \quad (11)$$

In both expressions, there are three possible outcomes. If neither agent is disciplined ($\theta \geqslant \phi_{0|1}$), then agent i enjoys benefit x_1^i now and infers the principal's type to be at least $\phi_{0|1}$. The agent then chooses $x_2^i = \phi_{0|1}$ in period 2, based on Lemma 1. If only one is disciplined ($\phi_{1|2} \leqslant \theta < \phi_{0|1}$), the leading agent suffers the penalty while the trailing agent remains safe. Both agents infer the principal's type to be at least $\phi_{1|2}$ and choose $x_2^i, x_2^{-i} = \phi_{1|2}$ in period 2. Finally, if both are disciplined ($\theta < \phi_{1|2}$), then each agent suffers the penalty and infers the principal's type to be below $\phi_{1|2}$. Both agents then choose $x_2^i, x_2^{-i} = 0$ to be safe. The positive enforcement externalities, which exist outside the proximity cone, are evident in (10). Observe that lower peer misconduct x_1^{-i} exposes agent i to a higher probability of enforcement; that is, the probability $F(\phi_{0|1}) = F(\frac{x_1^i(1-2\delta_P) - x_1^{-i}(2\delta_P)}{(1-2\delta_P)^2})$ is higher when x_1^{-i}

[5] While Figure 5 shows externalities when fixing a type of the principal, the analysis here takes expectation over the principal's types.

decreases.[6] Accordingly, to reduce his probability of enforcement, agent i is motivated to reduce his own misconduct to shield behind his peer.

4.3 Equilibrium

We now solve for the equilibrium agent misconduct in the main model that permits peer learning. We then compare the equilibrium misconduct in the main model to that in the opaque benchmark to show the effect of peer learning. The next theorem establishes our main result — the level of misconduct in the main model can either increase or decrease relative to the opaque benchmark depending on the principal's level of patience.

Theorem 1 (Effect of Peer Learning)

Compare to the opaque benchmark in which there is no peer learning:

(i) *For an impatient principal ($\delta_P < \frac{1}{2}$), when L is large enough, introducing peer learning leads to strictly* ***lower*** *expected misconduct in both periods and* ***lower*** *expected total discounted cost $\mathbb{E}(C)$ for all principal types.*[7]

(ii) *For a patient principal ($\delta_P \geqslant \frac{1}{2}$), introducing peer learning leads to strictly* ***higher*** *expected misconduct in both periods and* ***higher*** *expected total discounted cost $\mathbb{E}(C)$ for all principal types.*

Part (i) shows that, when the principal is sufficiently patient, peer learning reduces equilibrium misconduct, which is consistent with the established understanding in the collective experimentation literature. The literature suggests that allowing agents to learn from each other's experimentation can lead to a decline in experimentation due to free-riding (Bolton and Harris (1999); Keller, Rady, and Cripps (2005); Bonatti and Hörner (2011);

[6] According to (10), the probability of agent i being punished with penalty L is sum of the probability of only the leading agent being punished, $F(\phi_{0|1}) - F(\phi_{1|2})$, and the probability of both agents being punished, $F(\phi_{1|2})$; in aggregate, agent i's probability of enforcement equals $F(\phi_{0|1})$.

[7] We require a sufficiently large L to ensure tractability. We show numerically in Section 5.1 that the qualitative results remain unchanged if we lift this restriction. In the more general case, the period-1 equilibrium misconduct levels are not necessarily zero, but they continue to be low.

Chen (2020)). In contrast, part (ii) reveals that peer learning can actually induce equilibrium misconduct when the principal exhibits sufficient patience. This result is surprising for two reasons. First, the increase in misconduct (or experimentation) is typically not expected when agents can free-ride on each other's efforts to learn. Second, the unintended rise in misconduct occurs exactly when the principal is relatively patient and eager to suppress misconduct to establish a reputation for tough enforcement.

We explore this result in the next two propositions that characterize the agents' equilibrium misconduct when facing an impatient principal and a patient one, respectively.

Proposition 3 (Equilibrium Misconduct: $\delta_P < \frac{1}{2}$)
For every $\delta_P < \frac{1}{2}$, there exists a threshold $\hat{L}(\delta_P) < \frac{1+\delta-\delta_P}{h(0)}$ such that when $L > \hat{L}(\delta_P)$, there exists a unique equilibrium in pure strategies in which neither agent engages in misconduct in either period, i.e., $x_t^i = 0$, for $i = A, B$ and $t = 1, 2$.[8]

In contrast to the benchmark result in Proposition 1 where the period-1 equilibrium misconduct level is positive, Proposition 3 shows that peer learning causes the agents to eschew any misconduct when the principal is impatient.

This deterrence effect arises from two factors. First, there exist *positive* information externalities between the agents, which discourage misconduct as agents free-ride on each other's experimentation. Second, the opportunity to lower one's misconduct and shield behind his peer creates endogenous *positive* enforcement externalities, which further reduce the agents' misconduct. To understand the enforcement externalities, observe from Figure 3 that when $\delta_P < \frac{1}{2}$, the proximity cone does not cover the entire quadrant. Thus, given his peer's misconduct, an agent is able to determine whether the profile (x_1^A, x_1^B) stays within or outside of the proximity cone by adjusting his own level of misconduct. Importantly, this means that an agent can always choose to significantly reduce his own misconduct level to the point where he falls completely outside of the proximity cone. This allows him to enjoy

[8] The condition $L > \hat{L}(\delta_P)$ is to ensure tractability. We show numerically in Section 5.1 that the qualitative results remain unchanged if we lift this restriction. In the more general case, the period-1 equilibrium misconduct levels are not necessarily zero, but they continue to be low.

the positive information externalities his peer creates and avoid punishment. Indeed, the temptation to hide behind one's peer is irresistible, inducing both agents to reduce their misconduct. In equilibrium, neither agent engages in misconduct.

Now suppose the principal is relatively patient, that is, $\delta_P \geqslant \frac{1}{2}$.

Proposition 4 (Equilibrium Misconduct: $\delta_P \geqslant \frac{1}{2}$)
For $\delta_P \geqslant \frac{1}{2}$, there exists a unique equilibrium in pure strategies in which the agents choose the same misconduct level $x_1^A = x_1^B = x^{**}$ in period 1. The level x^{**} is greater than x^* in Proposition 1, and it uniquely solves:

$$\big(L + (1 + \delta - \delta_P)\phi_{0|2}(x^{**}, x^{**})\big)\, h\,\big(\phi_{0|2}(x^{**}, x^{**})\big) = 2 + \delta - 2\delta_P. \tag{12}$$

In addition, agent i's period-2 misconduct is $x_2^i = 0$ if $(e_1^A, e_1^B) = (1, 1)$, and $x_2^i = \phi_{0|2}(x^{**}, x^{**})$ if $(e_1^A, e_1^B) = (0, 0)$.

Comparing Proposition 4 to Proposition 1 reveals that peer learning actually *promotes* misconduct rather than suppressing it. When the principal is relatively patient, peer learning limits the principal's enforcement strategy to punishing both agents collectively, effectively tying the principal's hands; this enforcement strategy is illustrated in Figure 4(b), where the proximity cone extends to cover the entire quadrant. Intuitively, a more patient principal is concerned with her reputation in period 2 and therefore is hesitant to punish only one agent in period 1, as this could leak too much information — revealing her type to be within a narrow range. This enforcement strategy creates *negative* enforcement externalities between the agents, as one agent engages in higher levels of misconduct, knowing that part of the increased enforcement risk is shifted to his peer. Consequently, peer learning increases misconduct.

The negative enforcement externalities promote misconduct and work *against* the positive information externalities. In fact, Proposition 4 implies that the negative enforcement externalities encourage misconduct *more* than the positive information externalities discourage it, with the overall effect being heightened misconduct.

To understand why negative enforcement externalities dominate in equilibrium, suppose both agents start with the misconduct level in the opaque benchmark, i.e., $x_1^A = x_1^B = x^* > 0$. In the benchmark, an agent has no incentive to deviate from the equilibrium level x^* to a higher level, $x^* + \mathrm{d}x$, as the three forces are balanced against each other: the marginal risk of enforcement equals the sum of the marginal current private benefit of misconduct and the marginal future value from learning.

When information about the peer agent becomes transparent, however, two changes occur in the agent's marginal trade-offs. First, peer learning introduces positive information externalities, which reduce the marginal value of learning on one's own by half. This is because an agent can now free-ride on his peer's misconduct to learn, resulting in the value of learning being shared equally by both agents. Second, peer learning generates negative enforcement externalities, which also reduce the marginal enforcement risk by half. This is because the increase in the deviating agent's enforcement risk is shared equally between the two agents. As a result, two out of the three marginal forces are halved, while the remaining one, the marginal current private benefit, remains unaffected by peer learning. Therefore, the balance of forces in the opaque setting is broken, and an agent has an incentive to increase misconduct.

Lastly, we investigate the comparative statics on the patience of the principal, that is, how misconduct changes as δ_P varies. In the opaque benchmark, Corollary 1 establishes that a more patient principal (a higher δ_P) is always more capable of deterring misconduct, because she is more eager to signal having a low enforcement cost by engaging in more stringent enforcement in period 1. In the main model, where peer learning is present, however, misconduct is *non-monotonic* in the principal's patience δ_P, as the next corollary establishes.

Corollary 2 (Reputation Effect with Peer Learning)

Consider any L satisfying Assumption 1. For all $\delta_P < \frac{1}{2}$ such that $L > \hat{L}(\delta_P)$, the agents' misconduct in both periods remains zero. At $\delta_P = \frac{1}{2}$, the agents' misconduct in both periods *jumps up. For $\delta_P > \frac{1}{2}$, the agents' misconduct in both periods strictly decreases in δ_P.*

When there is peer learning, the principal's heightened reputation concern as δ_P increases has two countervailing effects. The first effect is that the principal implements a stricter enforcement criterion in period 1 to signal toughness, leading to a reduction in misconduct. This effect is also at play in the opaque benchmark. However, a second effect arises when the principal is more hesitant to punish a single agent due to her reputation concerns, and instead must punish both agents concurrently. This effect, when strong enough, prevents an agent from reducing his own misconduct to shield behind his peer. In such a scenario, both agents increase their misconduct to exploit the principal's vulnerability to maintain her reputation. As δ_P increases above 1/2, the second effect remains at its zenith while the first effect continues to grow, and accordingly, misconduct declines.

5 Extensions

We extend our analysis in several ways and show that the main qualitative results in Theorem 1 and Corollary 2 are unchanged.

5.1 Impatient Principal and Small Penalty

When there is peer learning and the principal is impatient, we find a unique symmetric equilibrium with $x_1^A = x_1^B = 0$ provided the penalty L is high enough, i.e., $L > \hat{L}(\delta_P)$ (see Proposition 3). In this section, we relax this assumption and allow the penalty to be small (i.e., $L \leqslant \hat{L}(\delta_P)$). We show that although the equilibrium is no longer symmetric or unique, we continue to find our main result that peer learning deters misconduct and reduces the principal's total discounted costs as in part (i) of Theorem 1. In addition, the comparative statics regarding δ_P in Corollary 2 continue to hold.

To show that peer learning reduces misconduct, let $\delta = 1$, $\delta_P = 0.1$, and the prior belief about the principal's type follows the exponential distribution with a c.d.f. of $F(\theta) =$

$1 - e^{-\theta}$. Assumption 1 requires $L \in [1, 1.9)$. The unique period-1 equilibrium is $x_1^A = x_1^B = 0$ when $L > \hat{L}(\delta_P) = 1.8$. For a smaller L, for example $L = 1.6$, there exist two asymmetric equilibria, where one agent chooses zero misconduct and the other engages in positive misconduct, specifically $(x_1^i, x_1^{-i}) = (0.09, 0)$, $i = A, B$. The opaque benchmark yields period-1 misconduct of $x_1^A = x_1^B = 0.14$, which is higher than both misconduct levels when peer learning is present. For an even smaller L, for example $L = 1$, there remain two asymmetric equilibria where both agents engage in positive misconduct of $(x_1^i, x_1^{-i}) = (0.37, 0.02)$. Observe that when L is sufficiently low, symmetry is broken as one of the two agents becomes the pioneer in learning while the other shields behind him. That said, the comparison with the opaque benchmark still holds: specifically, the period-1 misconduct is $x_1^A = x_1^B = 0.43$ in the benchmark compared to $(0.37, 0.02)$ when peer learning is present. Therefore, an impatient principal invokes transparency to deter misconduct and lower total discounted costs, as in the main model.

In addition, the non-monotonicity property in Corollary 2 also holds, implying that misconduct can be higher when the principal is more patient. Fixing $L = 1$, the expected misconduct in period 1 starts with $\frac{x_1^A + x_1^B}{2} = 0.38$ at $\delta_P = 0$, monotonically declines to zero as δ_P increases to $1/2$, jumps up to 0.33 at $\delta_P = 1/2$, and then declines again for $\delta_P > 1/2$.

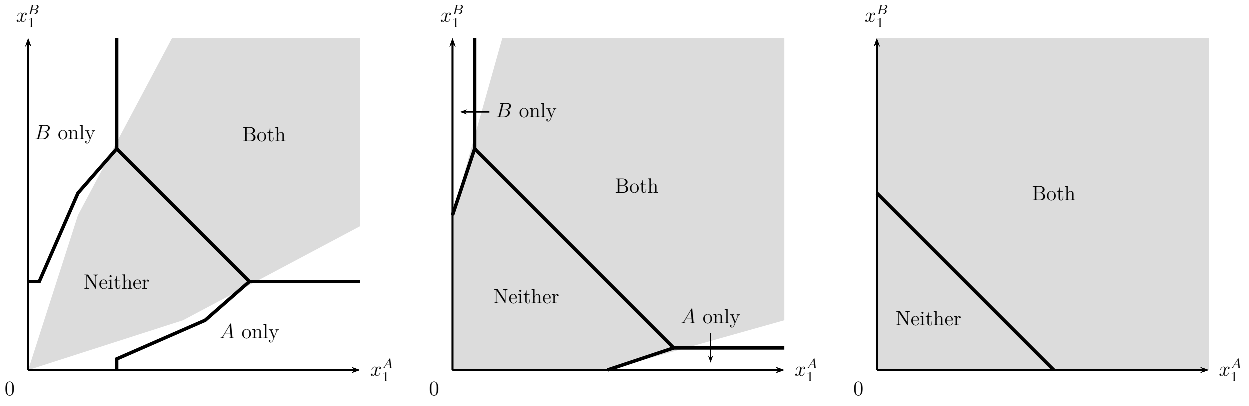
5.2 Interior Period-2 Misconduct

The main model assumes the penalty L cannot be too small (see Assumption 1(ii)) to ensure an agent choose the highest level of period-2 misconduct that avoids being disciplined, which equals the lower bound of the support of the updated beliefs about the principal's type. In this extension, we relax the restriction by allowing $Lh(0) < 1$. Intuitively, when the penalty L is not intimidating, an agent may bear some enforcement risk in period 2 by choosing misconduct in the interior of the support of the updated beliefs. The next lemma specifies the interior solution.

Lemma 3 *In period 2, if the posterior belief of agent i is the prior with truncated support $[\theta_L, \theta_H]$, then $x_2^i = x_2^*(\theta_L, \theta_H) \in [\theta_L, \theta_H)$ is the unique solution to $(L+x_2^i)h(x_2^i) = \frac{F(\theta_H)-F(x_2^i)}{1-F(x_2^i)}$ if $(L+\theta_L)h(\theta_L) < \frac{F(\theta_H)-F(\theta_L)}{1-F(\theta_L)}$ and $x_2^i = \theta_L$ otherwise.*

Lemma 3 shows that a weak penalty L can result in period-2 misconduct levels being higher than the lower bound of the truncated belief support. With weak deterrence, the principal's reputation boost from enforcement is less valuable, and intuitively, she is less willing to discipline the agents in period 1. Analogous to the main model, we can derive the principal's enforcement strategy in period 1. The properties of the principal's period-1 enforcement strategy, as characterized in Lemma 2, continue to hold because they do not rely on the strategies of the agents in period 2.

As the analytical form of the principal's strategy is convoluted, we offer the following numerical example to show the resemblance to the main model. Let $\delta = 1$, $L = 1$, and suppose $F(\theta) = 1 - e^{-\theta/2}$ is an exponential distribution of principal types. As $Lh(0) = 1/2 < 1$, Assumption 1 is violated. Figure 6 shows the enforcement zones in period 1 for a principal with type $\theta = 2$. When the principal is impatient ($\delta_P = 1/3$), the four enforcement zones resemble those in the main model, although the boundaries of the proximity cone are no longer linear. When $\delta_P = 1/2$, the principal tolerates misconduct by the less nefarious agent because reputation building is more difficult when the reputation is less valuable. When the principal is even more patient ($\delta_P = 2/3$), the enforcement zones align with Figure 4(a) in the main model. The resemblance generates similar incentives for the agents in period 1. For $\delta_P = 1/3$, a pure strategy equilibrium does not exist, but a mixed strategy equilibrium exists in which $x_1^A = 0.42$ with probability 0.76 and $x_1^A = 0.55$ with probability 0.24, while $x_1^B = 0.02$. This strategy profile yields an expected period-1 misconduct of 0.23, lower than 1.03 in the opaque benchmark, which is consistent with the result in part (i) of Theorem 1. For $\delta_P = 2/3$, the unique pure strategy equilibrium features $x_1^A = x_1^B = 0.89$, higher than 0.55 in the benchmark, which is consistent with the results in part (ii) of Theorem 1. Finally, the comparative statics for δ_P are similar to those in Corollary 2: the expected period-1



(a) $\delta_P = 1/3$ (b) $\delta_P = 1/2$ (c) $\delta_P = 2/3$

Figure 6: Principal's enforcement zones in period 1, when $Lh(0) = \frac{1}{2} < 1$. Grey areas show proximity sets. Parameters: $L = 1$, $F(\theta) = 1 - e^{-\theta/2}$, $\theta = 2$

misconduct is 0.23 at $\delta_P = 1/3$, declines to 0 as δ_P increases to 1/2, jumps to 1.27 at $\delta_P = 1/2$, and then declines again to 0.89 at $\delta_P = 2/3$.

5.3 Asymmetric Agents

So far we have assumed the agents are symmetric in the model's parameters. In this extension, we consider asymmetric agents. First, agents may suffer different penalty sizes. Second, with the same level of misconduct, agents may derive different private benefits. Third, agents may inflict different damages on the principal despite the same level of misconduct. With proper normalization, however, these asymmetries can be reduced into a difference in the agents' penalties.[9] Assume agent $i = A, B$ suffers a penalty L^i if punished, where both L^A and L^B satisfy Assumption 1. The modified penalty structure only affects the agents' misconduct in period 1, but not the principal's enforcement strategy. When characterizing the equilibrium, we replace the penalty L in agent i's payoff (9)-(11) with L^i.

Consider an impatient principal. Numerically, let $\delta = 1$, $\delta_P = 0.3$, and $F(\theta) = 1 - e^{-\theta}$. Assumption 1 requires $L^i \in [1, 1.7)$. If $(L^A, L^B) = (1.2, 1.5)$, then in the opaque benchmark

[9] In the second case, it is the ratio of private benefit to the cost of the penalty that matters for an agent's decision. Therefore, we can proportionally modify the parameters such that the private benefits equalize while the penalties may differ. In the third case, we can redefine misconduct as the damage to the principal, thereby reducing this case to the second case.

where peer learning is absent, the agents play individual games with the principal, and they choose period-1 misconduct of $x_1^A = 0.21$ and $x_1^B = 0.08$. In contrast, with peer learning, in period 1, there is a unique equilibrium in pure strategies in which $x_1^A = 0.06$ and $x_1^B = 0$. These levels of misconduct are both lower than those in the benchmark, which is consistent with Theorem 1(i).

Alternatively, consider a patient principal. Let $\delta = 1$, $\delta_P = 0.67$, and $F(\theta) = 1 - e^{-\theta}$. Assumption 1 requires $L^i \in [1, 1.33)$. If $(L^A, L^B) = (1, 1.25)$, then in the opaque benchmark, the agents play individual games with the principal and choose period-1 misconduct of $x_1^A = 0.08$ and $x_1^B = 0.02$. In contrast, with peer learning, in period 1 there is a unique equilibrium in pure strategies in which $x_1^A = 0.26$ and $x_1^B = 0.01$. The average period-1 misconduct between the two agents is 0.14, which is higher than the average misconduct in the opaque benchmark. The higher average misconduct with peer learning echos Theorem 1(ii).[10]

The non-monotonic effect of a change in the principal's patience as stated in Corollary 2 still applies in this asymmetric case. Fixing $\delta = 1$, $F(\theta) = 1 - e^{-\theta}$, $(L^A, L^B) = (1.2, 1.5)$, when $\delta_P = 0.3$, there is an average misconduct of 0.03 in period 1. It decreases to zero as δ_P increases to 1/2, jumps to 0.29 at $\delta_P = 1/2$, and decreases again as δ_P increases above 1/2.

5.4 Three Periods

In this section, we extend the horizon of the game from two to three periods. Applying the insights from the two-period model, we can conjecture the properties of equilibrium for a three-period model. In period 1, given misconduct (x_1^A, x_1^B), the principal's enforcement decision separates her types into two or three intervals, depending on her level of patience. In period 2, the agents update their beliefs to the corresponding interval and choose (x_2^A, x_2^B),

[10] With asymmetric agents, however, individual misconduct levels do not necessarily rise with peer learning. In the numerical example, while the average misconduct in period 1 increases due to peer learning, agent B's misconduct actually drops from 0.02 to 0.01. This is due to an additional force at work: the two agents' misconduct are strategic substitutes inside the proximity cone. Intuitively, agent A's penalty L^A is smaller than agent B's, so agent A is expected to have higher misconduct. This heightens agent B's enforcement risk and therefore reduces his misconduct. In sum, while peer learning increases the average misconduct levels, it also increases the discrepancy between the two agents' misconduct levels. The latter result does not exist in the main model due to the symmetry between the agents.

and then the principal further separates her types within that updated interval. Period 3 is the terminal period. As the principal types are potentially separated into intervals in each period, the benchmark and the main model become unwieldy as the number of periods increases. Nonetheless, numerical solutions are obtainable for a three-period game.

To illustrate the properties of the three-period game, we use backward induction starting with period 3. Invoking Lemma 1, the agents choose the highest misconduct level that avoids punishment. Moving back to period 2, the principal and the agents face a two-period continuation game.

On the one hand, consider an impatient principal. Let $\delta = 1$, $\delta_P = 0.3$, and $F(\theta) = 1 - e^{-\theta}$. We let $L = 1.5$, which satisfies Assumption 1. Absent peer-learning, the agents play individual games with the principal and choose period-1 misconduct of $x_1^i = 0.26$, which is adjusted down to $x_2^i = 0$ after being disciplined and up to $x_2^i = 0.43$ otherwise. There will be no enforcement in period 2, and the agents' misconduct in period 3 remains the same as in period 2. In contrast, with peer-learning, period 1 admits two asymmetric equilibria in pure strategies in which $x_1^i = 0.07$ and $x_1^{-i} = 0$. The principal never disciplines both agents. In period 2, both agents adjust misconduct to $x_2^i = 0$ if one agent is disciplined, and up to $x_2^i = 0.32$ if neither is disciplined. There will be no enforcement in period 2, and the misconduct in period 3 remains the same as in period 2. In sum, peer-learning reduces misconduct when the principal is impatient, which is consistent with part (i) of Theorem 1.

On the other hand, consider a patient principal. Let $\delta = 1$, $\delta_P = 0.6$, and $F(\theta) = 1 - e^{-\theta}$. Suppose $L = 1.2$, which still satisfies Assumption 1. Absent peer-learning, the agents play individual games with the principal and choose period-1 misconduct of $x_1^i = 0.02$, which is adjusted down to $x_2^i = 0$ after being punished and up to $x_2^i = 0.41$ otherwise. There will be no enforcement in period 2, and the misconduct in period 3 remains the same. In contrast, with peer-learning, there is a unique equilibrium in period 1 with $x_1^i = 0.02$. The principal never disciplines only one agent. In period 2, both agents adjust misconduct down to $x_2^i = 0$ after both are disciplined and up to $x_2^i = 0.43$ otherwise. There will be no enforcement in period 2, and the misconduct in period 3 remains the same. Accordingly, we observe that peer-learning heightens misconduct when the principal is patient, which is consistent with part (ii) of Theorem 1.

We also numerically confirm the counterpart of Corollary 2 in a three-period game. With three periods, the cutoff patience level for the principal is no longer $\delta_P = \frac{1}{2}$. Instead, the new proximity cone is defined by $\{(x_1^A, x_1^B) \in [0,\infty)^2 : x_1^B \geqslant x_1^A(1-2\delta_P(1+\delta_P)),\ x_1^A \geqslant x_1^B(1-2\delta_P(1+\delta_P))\}$, which implies the cone expands to the entire quadrant if and only if $\delta_P \geqslant \frac{\sqrt{3}-1}{2} = 0.37$. Fixing $\delta = 1$, $F(\theta) = 1 - e^{-\theta}$, $L = 1.5$, and $\delta_P = 0.3$, average misconduct is 0.04 in period 1. It approaches zero as δ_P increases to 0.37, before jumping up to 0.3 at $\delta_P = 0.37$, and decreases as δ_P continues to increase.

6 Conclusion

We examine a two-period model in which two agents make misconduct decisions in the presence of a strategic principal whose tolerance for misconduct is unknown to the agents. We depart from the existing literature on collective experimentation by allowing the principal – i.e., the object about which the players are learning – to be *strategic*. We show that enforcement externalities arise endogenously. As a result, the transparency of misconduct and enforcement, which allows the principal to "make an example" of the nefarious agent by publicly punishing the agent, may actually increase misconduct. We also show that reputation may benefit or hurt the principal, depending on the transparency of misconduct and enforcement. Our findings have implications for various institutional settings, including relations between headquarters and division managers, common owners and portfolio firm managers, and regulators and firms.

Appendix: Proofs

Proof of Lemma 1

Proof. (i) Since period 2 is the terminal period, the principal minimizes the continuation cost $\sum_{i=A,B}[x_2^i + e_2^i(\theta - x_2^i)]$. Therefore, $e_2^i = 1$ ($e_2^i = 0$) if $x_2^i > \theta$ ($x_2^i < \theta$). When $x_2^i = \theta$, Assumption 2 requires that $e_2^i = 0$.

(ii) In period 2, an agent i maximizes his continuation payoff $\mathbb{E}\left[x_2^i(1-e_2^i) - Le_2^i|q^i\right] = x_2^i(1-\mathbb{E}[e_2^i|q^i]) - L\mathbb{E}[e_2^i|q^i]$. According to the principal's strategy, $\mathbb{E}[e_2^i|q^i] = \Pr(x_2^i > \theta|q^i)$. If an agent's posterior belief is the prior truncated to $[\theta_L, \theta_H)$, then agent i, who chooses misconduct $x_2^i \in [\theta_L, \theta_H)$, receives a continuation payoff

$$x_2^i \frac{F(\theta_H) - F(x_2^i)}{F(\theta_H) - F(\theta_L)} - L\frac{F(x_2^i) - F(\theta_L)}{F(\theta_H) - F(\theta_L)}.$$

Differentiating w.r.t. x_2^i: $\frac{f(x_2^i)}{F(\theta_H)-F(\theta_L)}\left(\frac{F(\theta_H)-F(x_2^i)}{f(x_2^i)} - (L + x_2^i)\right) \leqslant \frac{f(x_2^i)}{F(\theta_H)-F(\theta_L)}\left(\frac{1}{h(0)} - L\right) < 0$, where the first inequality is due to Assumption 1 (i) and the second inequality is due to Assumption 1 (ii). This means the continuation payoff is strictly decreasing in $x_2^i \in [\theta_L, \theta_H)$. If $x_2^i < \theta_L$, then the payoff is x_2^i. If $x_2^i \geqslant \theta_H$, the agent payoff is $-L$. Therefore, the globally optimal choice is $x_2^i = \theta_L$. ■

Proof of Proposition 1

Proof. Given x_1^i, the principal's cost (2) from agent i, as a continuous function of θ, has a derivative of at least 1 almost everywhere. In contrast, the cost (3) has a derivative of at most $\delta_P < 1$ almost everywhere. Moreover, as $\theta \to \infty$, (2) is greater than (3). Therefore, the principal's strategy in period 1 must feature a cutoff $\theta^\dagger \geqslant 0$ such that $e_1^i = 1$ if and only if $\theta < \theta^\dagger$. The indifference is broken in favor of not disciplining because of Assumption 2. Then, according to Bayes' rule, agent i's posterior belief is the prior distribution truncated to $[\theta^\dagger, \infty)$ if $e_1^i = 0$, and truncated to $[0, \theta^\dagger)$ if $e_1^i = 1$. By Lemma 1, agent i chooses $x_2^i = \theta^\dagger$

and $x_2^i = 0$, respectively. Plugging into (2) and (3) and observe that (2) and (3) are equal at the cutoff type $\theta = \theta^\dagger$, we have $\theta^\dagger + \delta_P \cdot 0 = x_1^i + \delta_P \cdot \theta^\dagger$. Solving the equation, we have $\theta^\dagger = \phi(x_1^i) = \frac{x_1^i}{1-\delta_P}$.

In period 1, agent i's payoff is $(x_1^i + \delta\phi(x_1^i))\,(1 - F\,(\phi(x_1^i))) + (-L + \delta \cdot 0)F\,(\phi(x_1^i))$, its derivative having the same sign as $(1 + \delta - \delta_P) - (L + (1 + \delta - \delta_P)\phi(x_1^i))\,h\,(\phi(x_1^i))$. It is strictly decreasing in x_1^i by Assumption 1. It diverges to $-\infty$ as $x_1^i \to \infty$ and evaluates to $1 + \delta - \delta_P - Lh(0) > 0$ when $x_1^i = 0$. By Intermediate value theorem, there exists a unique $x^* > 0$ such that $(L + (1 + \delta - \delta_P)\phi(x_1^i))\,h\,(\phi(x_1^i)) = 1 + \delta - \delta_P$, and agent i's payoff is maximized at $x_1^i = x^*$. In addition, x^* is global optimal because the first derivative is positive when $x < x^*$ and negative when $x > x^*$. ■

Proof of Corollary 1

Proof. Equation (7) can be rewritten as

$$G(\delta_P, \phi(x^*)) \equiv (L + (1 + \delta - \delta_P)\phi(x^*))\,h\,(\phi(x^*)) - (1 + \delta - \delta_P) = 0.$$

Notice that $\frac{\partial G}{\partial \phi(x^*)} > 0$, and $\frac{\partial G}{\partial \delta_P} = 1 - \phi(x^*)h(\phi(x^*)) = \frac{Lh(\phi(x^*))}{1+\delta-\delta_P} > 0$ with the second equality being guaranteed by $G(\delta_P, \phi(x^*)) = 0$. By the Implicit function theorem, $\frac{\mathrm{d}\phi(x^*)}{\mathrm{d}\delta_P} = -\frac{\partial G}{\partial \delta_P} / \frac{\partial G}{\partial \phi(x^*)} < 0$. Since $x^* = \phi(x^*)(1 - \delta_P)$, we have x^* strictly decreases in δ_P. ■

Proof of Lemma 2

Proof. (i) Suppose for some principal type θ, we have $x_1^i > x_1^j$, $e_1^i = 0$ and $e_1^j = 1$. Let $\theta_L \leqslant \theta$ be the infimum of such types. The principal's total cost is at least $\theta + x_1^i + \delta_P(2\theta_L)$. By deviating to $e_1^i = 1$ and $e_1^j = 0$, the total cost is at most $\theta + x_1^j + \delta_P(2\theta)$. As $\theta \to \theta_L$, this deviation strictly lowers cost, a contradiction.

(ii) Depending on the principal's enforcement in period 1, agent i chooses potentially random misconduct level x_2^i in period 2 according to its posterior. Assume $x_1^A \geqslant x_1^B$ w.l.o.g.

By disciplining both agents, one agent, and neither agent in period 1, the principal respectively incurs the following total costs as continuous functions of θ

$$\begin{aligned}
C_2(\theta) &\equiv 2\theta + \delta_P \sum_i \mathbb{E}\left[\min\left\{\theta, x_2^i(x_1^A, x_1^B, e_1^A = 1, e_1^B = 1)\right\} \middle| \theta\right], \\
C_1(\theta) &\equiv \theta + x_1^B + \delta_P \sum_i \mathbb{E}\left[\min\left\{\theta, x_2^i(x_1^A, x_1^B, e_1^A = 1, e_1^B = 0)\right\} \middle| \theta\right], \\
C_0(\theta) &\equiv x_1^A + x_1^B + \delta_P \sum_i \mathbb{E}\left[\min\left\{\theta, x_2^i(x_1^A, x_1^B, e_1^A = 0, e_1^B = 0)\right\} \middle| \theta\right].
\end{aligned}$$

Define $\theta^* \equiv \inf\{\theta \geqslant 0 : C_0(\tilde{\theta}) \leqslant \min\{C_1(\tilde{\theta}), C_2(\tilde{\theta})\},\ \forall\ \tilde{\theta} > \theta\}$ and $\theta^{**} \equiv \sup\{\theta \geqslant 0 : C_2(\tilde{\theta}) < \min\{C_0(\tilde{\theta}), C_1(\tilde{\theta})\},\ \forall\ 0 \leqslant \tilde{\theta} < \theta\}$. By definition, $\theta^{**} \geqslant 0$. Since $C_2(\theta) > C_1(\theta) > C_0(\theta)$ as $\theta \to \infty$, we know $\theta^* < \infty$ and $\Pr\{\theta : (e_1^A, e_1^B) = (0,0)\} > 0$.

We prove a single-crossing property such that if $C_1(\theta_0) - C_0(\theta_0) \geqslant 0$ for some θ_0, then it remains true for all $\theta > \theta_0$. Suppose, towards contradiction, that $C_1(\theta_0) - C_0(\theta_0) \geqslant 0$ and $C_1(\theta_1) - C_0(\theta_1) < 0$ for some $\theta_1 > \theta_0$. Since $\lim_{\theta\to\infty} C_1(\theta) - C_0(\theta) = \infty$, by continuity there exist $\theta' \in [\theta_0, \theta_1)$ and $\theta'' > \theta_1$ such that $C_1(\theta') - C_0(\theta') = C_1(\theta'') - C_0(\theta'') = 0$ and $C_1(\theta) - C_0(\theta) < 0$ for all $\theta \in (\theta', \theta'')$. This means a type $\theta \in (\theta', \theta'')$ never chooses $(e_1^A, e_1^B) = (0,0)$, but since $(e_1^A, e_1^B) = (0,0)$ is on the equilibrium path, the agents assign zero conditional probability to $\theta \in (\theta', \theta'')$ upon seeing $(e_1^A, e_1^B) = (0,0)$. As a result, upon seeing $(e_1^A, e_1^B) = (0,0)$, x_2^i is never in the interval (θ', θ''), and thus for all $\theta \in (\theta', \theta'')$

$$\begin{aligned}
&C_1'(\theta) - C_0'(\theta) \\
&= 1 + \sum_i \delta_P \left(\Pr(x_2^i(x_1^A, x_1^B, e_1^A = 1, e_1^B = 0) \geqslant \theta \middle| \theta) - \Pr(x_2^i(x_1^A, x_1^B, e_1^A = 0, e_1^B = 0) \geqslant \theta \middle| \theta)\right) \\
&= 1 + \sum_i \delta_P \left(\Pr(x_2^i(x_1^A, x_1^B, e_1^A = 1, e_1^B = 0) \geqslant \theta \middle| \theta) - \Pr(x_2^i(x_1^A, x_1^B, e_1^A = 0, e_1^B = 0) \geqslant \theta'')\right),
\end{aligned}$$

which decreases in θ wherever defined. This implies $C_1(\theta) - C_0(\theta)$ is concave on $\theta \in (\theta', \theta'')$, a contradiction with the definition of θ' and θ''. Similarly, one can show that the function $C_2(\theta) - C_0(\theta)$ has the same single-crossing property.

There are two possibilities regarding $C_1(\theta)$. First, consider the case where $\{\theta : (e_1^A, e_1^B) =$

$(1,0)\} \neq \emptyset$. The above argument works for the function $C_2(\theta) - C_1(\theta)$ to show the single-crossing property. Then by definition of θ^* and θ^{**} and the three single-crossing properties, we have the cutoff strategy described in the lemma. Also, since $\{\theta : (e_1^A, e_1^B) = (1,0)\} = [\theta^{**}, \theta^*) \neq \emptyset$, we must have $\theta^{**} < \theta^*$.

Second, consider the case where $\{\theta : (e_1^A, e_1^B) = (1,0)\} = \emptyset$. Then by definitions of θ^* and θ^{**} and the single-crossing property of $C_2(\theta) - C_0(\theta)$, we have the cutoff strategy described in the lemma. Also, since $\{\theta : (e_1^A, e_1^B) = (1,0)\} = \emptyset$, we must have $\theta^{**} = \theta^*$. ■

Proof of Proposition 2

Proof. W.l.o.g., suppose $x_1^A \geqslant x_1^B$. (i) Let $(x_1^A, x_1^B) \in \mathcal{P}(\delta_P)$. Suppose, towards contradiction, that $\theta^{**} < \theta^*$. According to Lemma 2, type-θ^{**} principal incurs cost $\theta^{**} + x_1^B + 2\delta_P\theta^{**}$ when punishing only agent A, cost $2\theta^{**}+2\delta_P\cdot 0$ when punishing both, and cost $x_1^A+x_1^B+2\delta_P\theta^{**}$ when punishing neither. By definition, we must have $\theta^{**} + x_1^B + 2\delta_P\theta^{**} = 2\theta^{**} + 2\delta_P \cdot 0 < x_1^A + x_1^B + 2\delta_P\theta^{**}$, which yields $x_1^A > \theta^{**}$ and $\theta^{**}(1 - 2\delta_P) = x_1^B$. If $\delta_P < \frac{1}{2}$, this means $(x_1^A, x_1^B) \notin \mathcal{P}(\delta_P)$, a contradiction. If $\delta_P = \frac{1}{2}$, then $x_1^B = 0$. The definition of θ^* requires $x_1^A = \theta^{**}$. Fixing any type $\theta \in (\theta^{**}, \theta^*)$, we know $e_1^A(x_1^A, 0) = 1$ and $e_1^B(x_1^A, 0) = 0$. However, since $\phi_{0|2}(x_1^A, 0) = x_1^A = \theta^{**} < \theta$, for sufficiently small $\varepsilon > 0$, we have $\phi_{0|2}(x_1^A, \varepsilon) < \theta$. Then, $e_1^A(x_1^A, \varepsilon) = e_1^B(x_1^A, \varepsilon) = 0$, violating Assumption 2. If $\delta_P > \frac{1}{2}$, then $x_1^B = 0$ and $\theta^{**} = 0$. The definition of θ^* requires $x_1^A + (2\delta_P - 1)\theta^* = 0$, which contradicts the assumption that $\theta^* > \theta^{**} \geqslant 0$. In all above cases, we conclude $\theta^{**} = \theta^*$. Then the indifference of type θ^* requires $2\theta^* + 2\delta_P \cdot 0 = x_1^A + x_1^B + 2\delta_P\theta^*$, i.e., $\theta^{**} = \theta^* = \frac{x_1^A + x_1^B}{2(1-\delta_P)} = \phi_{0|2}(x_1^A, x_1^B)$.

(ii) Let $(x_1^A, x_1^B) \notin \mathcal{P}(\delta_P)$, which requires $\delta_P < \frac{1}{2}$. Suppose towards contradiction that $\theta^{**} = \theta^*$. According to Lemma 2, type-θ^* principal incurs cost $2\theta^* + 2\delta_P \cdot 0$ when punishing both, and cost $x_1^A + x_1^B + 2\delta_P\theta^*$ when punishing neither. By definition, we must have $2\theta^* + 2\delta_P \cdot 0 = x_1^A + x_1^B + 2\delta_P\theta^*$, which yields $\theta^* = \phi_{0|2}(x_1^A, x_1^B)$. Instead, the principal incurs a cost of at most $\theta^* + x_1^B + 2\delta_P\theta^*$ when punishing only agent A, regardless of the agents' beliefs. Since $(\theta^* + x_1^B + 2\delta_P\theta^*) - (2\theta^* + 2\delta_P \cdot 0) = \frac{x_1^B - (1-2\delta_P)x_1^A}{2(1-\delta_P)} < 0$, we have a profitable deviation, contradicting the assumption $\theta^{**} = \theta^{*}$. Therefore, $\theta^{**} < \theta^{*}$. For type θ^{**} to be indifferent between punishing both agents and punishing agent A, we require $2\theta^{**} + 2\delta_P \cdot 0 = \theta^{**} + x_1^B + 2\delta_P\theta^{**}$, i.e., $\theta^{**} = \frac{x_1^B}{1-2\delta_P} = \phi_{1|2}(x_1^A, x_1^B)$. For type θ^{*} to be indifferent between punishing agent A and punishing neither, we need $\theta^{*} + x_1^B + 2\delta_P\theta^{**} = x_1^A + x_1^B + 2\delta_P\theta^{*}$, i.e., $\theta^{*} = \frac{x_1^A(1-2\delta_P) - x_1^B(2\delta_P)}{(1-2\delta_P)^2} = \phi_{0|1}(x_1^A, x_1^B)$. ■

Proof of Proposition 3

Proof. First, we show that when L is close enough to $\frac{1+\delta-\delta_P}{h(0)}$, a pair (x_1^A, x_1^B) in the interior of the proximity cone is not part of an equilibrium. Suppose, towards contradiction, that it is, then the FOC of (9) w.r.t. x_1^A and x_1^B yields $x_1^A = x_1^B = x^{**}$, where x^{**} satisfies (12). Then, agent i's payoff simplifies to $U\left(\frac{x^{**}}{1-\delta_P}, L\right)$, where $U(\theta, L) \equiv (1+\delta-\delta_P)\theta(1-F(\theta)) - LF(\theta)$. Since $\frac{\partial U(\theta,(1+\delta-\delta_P)/h(0))}{\partial\theta} = -(1-F(\theta))(1+\delta-\delta_P)\left(\frac{h(\theta)}{h(0)} - 1 + \theta h(\theta)\right) < 0$ for all $\theta > 0$, by the Mean value theorem, $U\left(\frac{x^{**}}{1-\delta_P}, \frac{1+\delta-\delta_P}{h(0)}\right) < U\left(0, \frac{1+\delta-\delta_P}{h(0)}\right) = 0$. Because (12) defines x^{**} as a continuous function of L, we know $U\left(\frac{x^{**}}{1-\delta_P}, L\right) < 0$ for $L \in \left(\tilde{L}(\delta_P), \frac{1+\delta-\delta_P}{h(0)}\right)$ for some $\tilde{L}(\delta_P) < \frac{1+\delta-\delta_P}{h(0)}$. However, by choosing $x_1^i = 0$, agent i can secure a non-negative payoff from (11), a profitable deviation.

Next, we show that equilibrium cannot feature $x_1^A \neq x_1^B$. Suppose, towards contradiction, that $x_1^A > x_1^B$ in equilibrium. Since (x_1^A, x_1^B) is not in the interior of the proximity cone, it must be $x_1^A \geqslant \frac{x_1^B}{1-2\delta_P}$. In this regime, (10)'s derivative w.r.t. x_1^A is

$$\frac{1-F\left(\phi_{0|1}\right)}{1-2\delta_P}\left[(1+\delta-2\delta_P) - \left(L + x_1^A + \delta(\phi_{0|1} - \phi_{1|2})\right)h\left(\phi_{0|1}\right)\right]. \tag{13}$$

Given the definition of $\phi_{0|1}$ and $\phi_{1|2}$, the expression in the brackets is decreasing in x_1^A. When evaluated at $x_1^A = \frac{x_1^B}{1-2\delta_P}$, the expression becomes $(1+\delta-2\delta_P) - \left(L + \frac{x_1^B}{1-2\delta_P}\right)h\left(\frac{x_1^B}{1-2\delta_P}\right) \leqslant 1+\delta-2\delta_P - Lh(0) < 0$ when $L > \frac{1+\delta-2\delta_P}{h(0)}$. That is, if $L > \frac{1+\delta-2\delta_P}{h(0)}$, agent A's optimization requires $x_1^A = \frac{x_1^B}{1-2\delta_P}$. To be consistent with an equilibrium, agent B must be willing to choose $x_1^B = x_1^A(1-2\delta_P)$. However, the derivative of (11) w.r.t. x_1^B, when evaluated at $x_1^B =$

$x_1^A(1-2\delta_P)$, is $-\frac{1-F(x_1^A)}{(1-2\delta_P)^2}((1+\delta-2\delta_P)(2\delta_P)+(1-2\delta_P)((L+(1+\delta-2\delta_P)x_1^A)h(x_1^A)-1)) < 0$, so that agent B will deviate to some lower x_1^B.

Finally, the only remaining possibility is $x_1^A = x_1^B = 0$, and we verify it as an equilibrium. Given $x_1^B = 0$, agent A's payoff is described by (10), with derivative described by (13). Since $x_1^B = 0$, we have $\phi_{0|1} = \frac{x_1^A}{1-2\delta_P}$ and $\phi_{1|2} = 0$. When $L = \frac{1+\delta-2\delta_P}{h(0)}$, (13) evaluated at $x_1^A = 0$ is zero. Therefore, for all larger L, $x_1^A = 0$ is the best response for agent A. The same applies to x_1^B.

For the above arguments to hold, we require $L > \hat{L}(\delta_P) \equiv \max\left\{\tilde{L}(\delta_P), \frac{1+\delta-2\delta_P}{h(0)}\right\}$. Since $x_1^i = 0$ in period 1, no agent is disciplined, and both choose zero misconduct in period 2. ■

Proof of Proposition 4

Proof. For $\delta_P \geqslant \frac{1}{2}$, any (x_1^A, x_1^B) is in the proximity cone. The FOC of (9) w.r.t. x_1^i requires

$$(2+\delta-2\delta_P) - \left(L + x_1^i + \delta\phi_{0|2}\right) h\left(\phi_{0|2}\right) \leqslant 0, \tag{14}$$

and equality holds if $x_1^i > 0$. The left-hand side of (14) is decreasing in x_1^i, and it diverges to $-\infty$ as $x_1^i \to \infty$. Therefore, given any $x_1^{-i} \geqslant 0$, there exists a unique best response $x_1^i(x_1^{-i}) \geqslant 0$ of agent i. If the best response $x_1^i(x_1^{-i}) > 0$, the Implicit function theorem implies that

$$\frac{\mathrm{d}x_1^i}{\mathrm{d}x_1^{-i}} = -1 + \frac{2(1-\delta_P)h\left(\phi_{0|2}\right)^2}{(2+\delta-2\delta_P)\left(h\left(\phi_{0|2}\right)^2 + h'\left(\phi_{0|2}\right)\right)} \in (-1, 0).$$

If the best response $x_1^i(x_1^{-i}) = 0$ for some x_1^{-i}, then as the left-hand side of (14) decreases in x_1^{-i}, we have $x_1^i(x_1^{-i}) = 0$ for all larger x_1^{-i}. Consequently, the best response functions intersect at most once, ensuring uniqueness of the equilibrium. Guessing $x_1^A, x_1^B > 0$, the FOCs hold with equality. Solving the system, we have $x_1^A = x_1^B \equiv x^{**}$, where x^{**} satisfies (12). The left-hand side of (12) strictly increases in x^{**}. At $x^{**} = 0$, the left-hand side becomes $Lh(0)$, which is small than the right-hand side $2+\delta-2\delta_P$ according to Assumption 1. As $x^{**} \to \infty$, the left-hand side diverges to $\infty > 2+\delta-2\delta_P$. Therefore, (12) has a unique solution $x^{**} > 0$ by the Intermediate value theorem, and our guess of the equilibrium is verified. The misconduct in period 2 follows from Lemma 1. ■

Proof of Theorem 1

Proof. (i) When $\delta_P < \frac{1}{2}$ and $L \in \left(\hat{L}(\delta_P), \frac{1+\delta-\delta_P}{h(0)}\right)$, the agents in the main model choose misconduct level 0 in period 1, and again 0 in period 2, according to Proposition 3. A type-θ principal's expected total cost is 0. In the benchmark, the agents choose misconduct level $x^* > 0$ in period 1 that satisfies (7), and they choose either $\phi(x^*)$ or 0 in period 2 depending on being disciplined or not. A type-θ principal's expected total discounted cost simplifies to $\min\{2\theta, 2\phi(x^*)\}$. Therefore, the expected misconduct is lower in both periods and the expected cost is lower when there is peer learning.

(ii) When $\delta_P \geqslant \frac{1}{2}$, the agents in the main model choose misconduct level x^{**} in period 1 that satisfies (12), according to Proposition 4. In period 2, misconduct is either $\phi_{0|2}(x^{**}, x^{**})$ or 0, depending on whether $\theta \geqslant \phi_{0|2}(x^{**}, x^{**})$ or not. The expected misconduct in period 2 is $\left(1 - F\left(\phi_{0|2}(x^{**}, x^{**})\right)\right) \phi_{0|2}(x^{**}, x^{**})$. A type-$\theta$ principal's expected total discounted cost simplifies to $\min\{2\theta, 2\phi_{0|2}(x^{**}, x^{**})\}$. In the benchmark, the agents choose misconduct level $x^* > 0$ in period 1 that satisfies (7), and they choose either $\phi(x^*)$ or 0 in period 2 depending on whether $\theta \geqslant \phi(x^*)$ or not. The expected misconduct in period 2 is $(1 - F(\phi(x^*)))\,\phi(x^*)$. A type-$\theta$ principal's expected total discounted cost is $\min\{2\theta, 2\phi(x^*)\}$.

Since $2+\delta-2\delta_P > 1+\delta-\delta_P$ and the function $(L + (1+\delta-\delta_P)\theta)\,h(\theta)$ is increasing in x, we know $\phi_{0|2}(x^{**}, x^{**}) > \phi(x^*)$, i.e., $x^{**} > x^*$, so that misconduct is higher in period 1 when enforcement is transparent. For period 2, notice that the function $(1 - F(\theta))\theta$ has derivative $(1-F(\theta))(1-\theta h(\theta))$. Since

$$\begin{aligned} 2+\delta-2\delta_P &= \left(L+(1+\delta-\delta_P)\phi_{0|2}(x^{**},x^{**})\right) h\left(\phi_{0|2}(x^{**},x^{**})\right) \\ &\geqslant Lh(0)+(1+\delta-\delta_P)\phi_{0|2}(x^{**},x^{**})h\left(\phi_{0|2}(x^{**},x^{**})\right) \\ &> 1-\delta_P+(1+\delta-\delta_P)\phi_{0|2}(x^{**},x^{**})h\left(\phi_{0|2}(x^{**},x^{**})\right), \end{aligned}$$

we know that $\phi_{0|2}(x^{**},x^{**})h\left(\phi_{0|2}(x^{**},x^{**})\right) < 1$, implying that the derivative $(1-F(\theta))(1-\theta h(\theta))$ is positive for $\theta = \phi_{0|2}(x^{**},x^{**})$. According to Assumption 1, $(1-F(\theta))(1-\theta h(\theta))$ is also positive all $\theta \leqslant \phi_{0|2}(x^{**},x^{**})$. Therefore, $(1-F(\theta))\,\theta$ is lower if evaluated at $\theta = \phi(x^*)$ than if evaluated at a higher level $\theta = \phi_{0|2}(x^{**},x^{**})$, and misconduct in period 2 is higher when there is peer learning. Also, since $\phi_{0|2}(x^{**},x^{**}) > \phi(x^*)$, the principal's cost is weakly higher with peer learning and strictly higher if $\theta > \phi(x^*)$. ■

Proof of Corollary 2

Proof. When $\delta_P < \frac{1}{2}$ and $L > \hat{L}(\delta_P)$, the conclusion follows from Proposition 3 and the proof of Theorem 1. When $\delta_P = \frac{1}{2}$, the conclusion follows from Proposition 4 and the proof of Theorem 1.

When $\delta_P > \frac{1}{2}$, (12) can be rewritten as

$$G(\delta_P, \phi_{0|2}(x^{**},x^{**})) \equiv (L+(1+\delta-\delta_P)\phi_{0|2}(x^{**},x^{**}))h(\phi_{0|2}(x^{**},x^{**})) - (2+\delta-2\delta_P) = 0.$$

Notice that G strictly increases in $\phi_{0|2}(x^{**},x^{**})$, and $\frac{\partial G}{\partial \delta_P} = 2-\phi_{0|2}(x^{**},x^{**})h(\phi_{0|2}(x^{**},x^{**})) = \frac{Lh(\phi_{0|2}(x^{**},x^{**}))+\delta}{1+\delta-\delta_P} > 0$. By the Implicit function theorem, $\phi_{0|2}(x^{**},x^{**})$ strictly decreases in δ_P. Therefore, x^{**}, which is equal to $\phi_{0|2}(x^{**},x^{**})(1-\delta_P)$, also strictly decreases in δ_P.

The expected misconduct in period 2 is $\left(1-F\left(\phi_{0|2}(x^{**},x^{**})\right)\right)\phi_{0|2}(x^{**},x^{**})$ according to the proof of Theorem 1. From $G(\delta_P, \phi_{0|2}(x^{**}, x^{**})) = 0$, we have

$$\phi_{0|2}(x^{**}, x^{**})h(\phi_{0|2}(x^{**}, x^{**})) = \frac{2+\delta-2\delta_P - Lh(\phi_{0|2}(x^{**}, x^{**}))}{1+\delta-\delta_P} < \frac{1+\delta-2\delta_P}{1+\delta-\delta_P} < 1.$$

Since the function $(1-F(\theta))\theta$ has derivative $(1-F(\theta))(1-\theta h(\theta))$, which is positive whenever $\theta \leqslant \phi_{0|2}(x^{**}, x^{**})$. Therefore, a higher δ_P, leading to a lower $\phi_{0|2}(x^{**}, x^{**})$, also lowers the expected misconduct in period 2. ■

Proof of Lemma 3

Proof. If the posterior belief is the prior truncated to $[\theta_L, \theta_H)$, then agent i, who chooses misconduct $x_2^i \in [\theta_L, \theta_H)$, receives a continuation payoff:

$$x_2^i \frac{F(\theta_H) - F(x_2^i)}{F(\theta_H) - F(\theta_L)} - L\frac{F(x_2^i) - F(\theta_L)}{F(\theta_H) - F(\theta_L)}.$$

Differentiating with respect to x_2^i yields $\frac{1-F(x_2^i)}{F(\theta_H)-F(\theta_L)} \left(\frac{F(\theta_H)-F(x_2^i)}{1-F(x_2^i)} - (L + x_2^i)h(x_2^i) \right)$. The Karush-Kuhn-Tucker conditions require $\frac{F(\theta_H)-F(x_2^i)}{1-F(x_2^i)} - (L + x_2^i)h(x_2^i) \leqslant 0$, with equality if $x_2^i = \theta_L$. Since the left-hand side is decreasing in x_2^i, this condition is also sufficient. ■

References

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991). Optimal learning by experimentation. *Review of Economic Studies 58*(4), 621–654.

Bar-Isaac, H. and J. Deb (2014). (Good and bad) reputation for a servant of two masters. *American Economic Journal: Microeconomics 6*(4), 293–325.

Bergemann, D. and U. Hege (2005). The financing of innovation: Learning and stopping. *RAND Journal of Economics*, 719–752.

Bergemann, D. and J. Välimäki (2000). Experimentation in markets. *Review of Economic Studies 67*(2), 213–234.

Bergemann, D. and J. Välimäki (2006). Bandit problems. Cowles Foundation discussion paper.

Bolton, P. and C. Harris (1999). Strategic experimentation. *Econometrica 67*(2), 349–374.

Bonatti, A. and J. Hörner (2011). Collaborating. *American Economic Review 101*, 632–663.

Bond, P. and K. Hagerty (2010). Preventing crime waves. *American Economic Journal: Microeconomics 2*(3), 138–59.

Boot, A. W. and A. V. Thakor (1993). Self-interested bank regulation. *American Economic Review 83*(2), 206–212.

Chen, Y. (2020). A revision game of experimentation on a common threshold. *Journal of Economic Theory 186*, 104997.

Corona, C. and R. S. Randhawa (2010). The auditor's slippery slope: An analysis of reputational incentives. *Management science 56*(6), 924–937.

Deb, R., M. Mitchell, and M. M. Pai (2022). (Bad) reputation in relational contracting. *Theoretical Economics 17*(2), 763–800.

Ely, J. C. and J. Välimäki (2003). Bad reputation. *The Quarterly Journal of Economics 118*(3), 785–814.

Goldstein, I. and Y. Leitner (2020). Stress tests disclosure: Theory, practice, and new perspectives. In J. D. Farmer, A. M. Kleinnijenhuis, T. Schuermann, and T. Wetzer (Eds.), *Handbook of Financial Stress Testing.* Cambridge University Press.

Goldstein, I. and H. Sapra (2013). Should banks' stress test results be disclosed? An analysis of the costs and benefits. *Foundations and Trends in Finance 8*(1), 1–54.

Hörner, J. and A. Skrzypacz (2016). Learning, experimentation and information design. In

Advances in Economics and Econometrics: Eleventh World Congress, Volume 1, pp. 63–98.

Huang, C. (2017). Defending against speculative attacks: The policy maker's reputation. *Journal of Economic Theory 171*, 1–34.

Keller, G., S. Rady, and M. Cripps (2005). Strategic experimentation with exponential bandits. *Econometrica 73*(1), 39–68.

Manso, G. (2011). Motivating innovation. *Journal of Finance 66*(5), 1823–1860.

Marinovic, I. and M. Szydlowski (2022). Monitoring with career concerns. *The RAND Journal of Economics 53*(2), 404–428.

Morrison, A. D. and L. White (2013). Reputational contagion and optimal regulatory forbearance. *Journal of Financial Economics 110*(3), 642–658.

Nanda, R. and M. Rhodes-Kropf (2017). Financing risk and innovation. *Management Science 63*(4), 901–918.

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *Journal of Economic Theory 9*(2), 185–202.

Shapiro, J. and D. Skeie (2015). Information management in banking crises. *Review of Financial Studies 28*(8), 2322–2363.