Firm Design when Monitors Can Be Bribed*

Manuel Lleonart-Anguix†

October 27, 2025

Latest version here.

Abstract

This paper develops a framework to design internal monitoring structures that minimize the cost of enforcing honesty. I study internal monitoring when supervisors themselves can be bribed. The firm jointly chooses the *architecture* of oversight (how many supervisors per worker and how their spans of control overlap) and the *incentives* needed to keep supervisors honest. Dual monitoring—assigning two independent supervisors to the same worker—makes wrongdoing harder to conceal and lowers the wage required to maintain honesty; the resulting wage savings can outweigh the extra hiring cost. The analysis shows that redundancy is optimal when monitoring technology is effective and collusion pressures are high—precisely the environments where overlap in monitoring is most valuable, that is, when verifiability is strong and monitors' outside options are weak. By linking organizational structure to the credibility of enforcement, the paper reframes dual monitoring not as waste but as a rational enforcement tool.

*I am grateful to Pau Milán for his invaluable guidance throughout the development of this paper. Part of this research was conducted during my visit to the Paris School of Economics, under the kind invitation of Francis Bloch, to whom I am also deeply indebted. I thank Catherine Bobtcheff, Mikhail Drugov, Inés Macho-Stadler, Antonine Macé, Paul-Henri Moisson, Fernando Payro, David Pérez-Castrillo, Patrick Sewell, Marta Troya, and seminar participants at the BSE Jamboree, the Microlab at UAB, and the ENTER Jamboree at the SSE for their helpful comments and suggestions.

†Universitat Autònoma de Barcelona and Barcelona School of Economics. Contact: manuel.lleonart@bse.eu

1 Introduction

In low- and middle-income countries, weak governance and internal oversight failures cost firms an estimated 2–5% of annual revenue—losses comparable to major tax burdens or trade barriers (ACFE, 2024; World Bank, 2006). Across industries and countries, the presence of multiple supervisors on the same task has typically been explained as a way to reduce mistakes, segregate duties, or combine different skills. In such accounts, duplication arises when tasks are complex or multidimensional—requiring error reduction (Sah and Stiglitz, 1986), segregation of duties (Kobelsky, 2014), and knowledge complementarities (Garicano, 2000; MacDuffie, 1995). These accounts do not predict *dual monitoring* for simple, standardized tasks—there redundancy is treated as waste—and they ignore the possibility that monitors can be bribed. Firms, regulators, and public agencies face this problem daily when they rely on auditors, inspectors, or recruiters who might collude with those they oversee. Modeling incentive pay and oversight structure jointly reveals a different logic. Assigning two independent supervisors to the same worker can be optimal not because the task is complex, but because redundancy lowers the pay needed to keep each supervisor honest. This shifts how we think about assigning supervisors across workers and when overlap should be part of efficient design.

This paper develops a framework to study the optimal design of internal oversight when supervisors themselves can be bribed. I jointly endogenize the monitoring architecture—how many supervisors oversee each worker and how their spans of control (the number of workers assigned to each supervisor) overlap—and the pay needed to keep supervisors honest. Here, overlap means assigning two independent supervisors to the same worker. By building bribery risk into the choice of structure and pay, the model shows that dual monitoring can reduce the wage required to maintain honesty; in some environments, the wage savings exceed the cost of hiring an additional supervisor. What looks like waste under standard theory can therefore be a rational response to corruption risk. The framework links oversight design to monitoring technology and incentive pay, and develops a practical mechanism to deter collusion: *dual monitoring* with aligned incentives that make bribery unattractive. It reframes effective oversight as disciplining monitors as well as supervising workers.[1]

The environment is a principal–supervisor–worker hierarchy in the spirit of Tirole (1986), with a continuum of workers performing identical tasks. Each worker is high- or low-productivity, with type privately known to them. The principal hires one or more supervisors

[1] Versions of overlapping oversight exist in practice—for example, teacher evaluations by school administrators and external reviewers (DCPS, 2023), dual authorizations in banking (BCBS, 2011; RBI, 2011), two-person custody rules (OCC, 2012), multiple signatories on reports (FDIC, 2014), and overlapping financial regulators (Keightley and Jickling, 2017); international guidelines also endorse "four-eyes" safeguards (OECD, 2018).

to oversee subsets of workers and report deviations. Detection is imperfect and declines with a supervisor's span of control (attention is scarce). Workers can bribe their monitors to suppress adverse reports. The firm chooses (i) how many monitors per worker and how spans overlap, and (ii) a wage policy that renders accepting bribes unattractive. Crucially, the frequency and profitability of bribe opportunities depend on the same choices that govern detection: structure and pay are inseparable design margins.

Two levers interact. First, architecture: assigning two independent supervisors to the same worker raises the cost of silence, since a deviator must bribe both. Second, pay: compensation makes acceptance of any bribe unattractive, accounting for how span affects detection and the flow of potential bribes. Compensation can also be tied to the other supervisor's reports, creating competition to report and further discouraging collusion. Because structure and pay co-determine both detection and bribe incentives, the monitoring unit operates as an endogenous enforcement technology rather than a fixed input.

First, the possibility of collusion overturns the one-monitor-per-worker rule: when baseline verifiability is high or detection decays gradually with span, the firm optimally assigns two independent monitors to the same worker. Second, the monitor's wage is non-monotonic in the span of control: with very small or very large spans, the expected number of detections—and hence the monitor's expected wage—is low, while at intermediate spans it peaks. Increasing overlap then reduces the honesty premium and can lower per-monitor pay even as the number of monitors rises, so the total wage bill need not increase—and may even fall. This creates a trade-off between paying more for loyalty and hiring more monitors to share oversight; better monitoring technology shifts the optimum toward increases collusion pressure and shifts the optimum toward narrower spans and more redundancy. Third, labor-market conditions shape the case for overlap: when supervisors have weak outside options, most of their pay is devoted to deterring bribery and overlap trims that premium, making dual supervision relatively cheap; when outside options are strong, much of pay is needed simply to retain supervisors, the savings from overlap are small, and the firm favors wider spans without duplication. These patterns link observable features of technology and governance with clear policies about hierarchy width, double monitoring and wage structure and allow me to measure the loses from naïve designs.

By embedding collusion risk into the design problem, the model reframes dual monitoring from bureaucratic slack into a practical mechanism to deter collusion. This makes firm design a tool against collusion: use dual monitoring where verifiability is strong and align incentives to eliminate the gains from bribery. The analysis also provides a basis for evaluating anti-corruption policies that strengthen internal governance—through redundancy, span limits, and incentive design—rather than relying solely on external enforcement. Finally, it provides simple design rules for when redundancy in oversight is efficient and when it can be safely reduced, and suggests empirical tests linking monitoring structures to enforcement strength across countries and sectors.

Literature Review. A central theme in the economics of organizations is that the architecture of hierarchies shapes how firms enforce discipline, process information, and allocate expertise. Three traditions have been particularly influential in shaping this view, each emphasizing a different function of hierarchy—monitoring, screening, or knowledge—and together they provide the intellectual backdrop for my analysis.

The first tradition views firms as *monitoring hierarchies*, where the key challenge is the *loss of control* as firms grow, so wages' increase need to compensate to keep workers from shirking (Williamson, 1967; Calvo and Wellisz, 1978; Calvo and Wellisz, 1979; Keren and Levhari, 1983; Qian, 1994). Later work ties these internal trade-offs to finance and market structure (Acemoglu and Newman, 2002; Chen, 2017; Chen and Suen, 2019). What these models typically rule out is that layers can "talk." If a monitor can collude with a worker, the principal can choose spans and whether to use *dual monitoring* (two independent monitors on the same worker) to deter collusion. In my framework, that choice matters because dual monitoring changes the pay needed to keep each monitor honest. The upshot is simple: when communication enables collusion, dual monitoring can be a cost-effective way to sustain discipline.

A second tradition treats hierarchies as *screening devices* that reduce errors in decision-making. Sah and Stiglitz (1986) showed that organizational form shapes the distribution of mistakes: polyarchies reduce false negatives, hierarchies reduce false positives. Ioannides (2012) extended this perspective, showing that in uncertain environments firms optimally introduce redundancy to protect against costly errors. This literature highlights how duplication and overlap can be safeguards even when they appear wasteful. My contribution reframes this mechanism: in my setting, redundancy is not only about error correction but also about raising the price of silence, transforming overlap into an anti-collusion device.

A third branch sees firms as *knowledge hierarchies*. In Garicano (2000), workers specialize in routine tasks while supervisors handle exceptional problems requiring expertise, and follow-up work shows how this structure generates both sorting and steep wage gradients (Garicano and Rossi-Hansberg, 2004; Garicano and Rossi-Hansberg, 2006). Here, hierarchy explains inequality and the allocation of knowledge. My model abstracts from heterogeneity in expertise, fixing the number of layers, and instead makes supervisors themselves strategic actors who may collude. The organizational form thus emerges not from knowledge allocation but from the need to sustain honest enforcement.

Taken together, these three literatures show how hierarchies matter for supervision, decision-making, and knowledge. Yet they largely abstract from the possibility that monitors themselves may be corruptible. Once this is recognized, oversight ceases to be a neutral technology and becomes an endogenous choice: firms may design spans, overlap, and wages not only to mitigate information frictions but also to deter collusion. This is the gap my analysis addresses.

Building on this foundation, a second set of contributions studies collusion and contracting more directly. In the classic principal–supervisor–agent framework, collusion undermines incentives and forces the principal to adapt contracts (Tirole, 1986; Kofman and Lawarrée, 1996; Rahman, 2012; Compte, 1995; Jean-Jacques Laffont and Martimort, 2000). This perspective naturally links to the theory of common agency: while Bernheim and Whinston (1986) modeled multiple principals bidding for influence, overlapping monitors create a mirror problem in which agents bid for silence. My framework formalizes this logic, making organizational architecture part of the tools that the principal can use to fight collusion. I build on a large *efficiency-wage* literature in which imperfect monitoring leads firms to pay above-market wages to deter shirking (Shapiro and Stiglitz, 1984). I add that rents are shaped not only by imperfect monitoring but also by potential collusion between monitors and workers, and I endogenize firm architecture—span and *dual monitoring*—to show when structure lowers the required honesty premia.

Finally, a large literature studies corruption and enforcement institutions. Classic work emphasized corruption as a distortionary tax or capture of regulators (Rose-Ackerman, 1978; Shleifer and Vishny, 1993; Jean-Jacques Laffont and Tirole, 1991), while recent contributions examine collusive networks inside organizations (Drugov, 2010; Fan, 2022). Empirical evidence confirms that oversight design matters: stronger auditor incentives reduce misreporting in India (Duflo et al., 2013), adding a second countersignature lowered procurement prices in Pakistan (Bandiera et al., 2021), and overlapping jurisdictions mitigate capture in administrative law (Gersen, 2007). My analysis draws on these insights by treating "four-eyes" policies as a firm-level analogue to overlapping jurisdictions: costly, but optimal when the risk of bribery is high.

In sum, the existing literatures explain how hierarchies shape incentives, information, and expertise; how contracts adapt to collusion; and how enforcement institutions can be corrupted or strengthened. My contribution is to bring these strands together, modeling the internal governance of firms as a corruption-control technology and showing how redundancy and overlap can sustain honest enforcement even when monitors themselves are at risk of capture.

Road map. The rest of the paper is organized as follows. Section 2 introduces the theoretical framework. Section 3 analyzes the benchmark without collusion. Section 4 studies optimal monitoring under collusion. Section 5 analyzes the loses from naïve results. Section 6 concludes.

2 Model

This paper develops a continuous version of the classic Principal–Supervisor–Agent framework introduced by Tirole (1986), extending it to allow the principal to endogenously choose the monitoring hierarchy within the firm. While detailed descriptions of each agent type follow, it suffices for now to view the principal as the residual claimant of firm profits, the agents as the sole productive units, and the monitors as intermediaries whose role is to report agents' productivity in order to mitigate the informational rents arising from asymmetric information.

The principal lacks both the time and the ability to directly observe agent behavior. Therefore, she can hire monitors to perform this task on her behalf.

Agents There is a continuum of agents, uniformly distributed over the interval $A = [0, 1]$. All agents are ex-ante identical. Let $i \in A$ denote a generic agent. Each agent generates output x_i given by:

$$x_i = \theta_i + e_i,$$

where θ_i represents the agent's productivity and $e_i \geq 0$ denotes effort. Productivity is binary: $\theta_i \in \{\theta_H, \theta_L\}$, with $\theta_H > \theta_L \geq 0$. The productivity type is drawn from a uniform distribution: $q \sim U(0, 1)$, implying that the expected share of high- and low-productivity agents is equal. Productivity realizations are independent across agents.

I define the skill diversity as $\Delta\theta = \theta_H - \theta_L \leq \frac{1}{2}$, such that high-productivity agents can mimic low-productivity ones by exerting strictly positive effort. This assumption is crucial for generating a non-trivial contracting problem under asymmetric information.[2]

Agents observe their productivity before choosing effort. Effort entails a cost given by $c(e_i) = \frac{e_i^2}{2}$. Each agent signs a contract with the principal specifying a wage $W(x_i, r_i)$,

[2] If instead $\Delta\theta > \frac{1}{2}$, mimicking would require high-productivity agents to exert negative effort, violating the constraint $e_i \geq 0$. One could alternatively model agents as being able to destroy output at a cost, i.e., allow $e_i < 0$, but this approach is not pursued here.

contingent on the realized output (observable to all parties) and a report r_i issued by the monitors (defined below). Agents are risk neutral, with utility $U_i(x) = x$, so expected utility is $\mathbb{E}[W(x_i, r_i) - c(e_i)]$. The participation constraint for each agent is:

$$\mathbb{E}[W(x_i, r_i) - c(e_i)] \geq 0. \tag{PCW}$$

Monitors The principal cannot directly observe either agents' effort or their productivity. To address this, she may hire monitors who are tasked with observing the productivity of a subset of agents. Let M denote the set of monitors. Each monitor $j \in M$ observes a subset $I_j \subseteq A$, with $\mu_j = \int_{I_j} di$ denoting monitor j's span of control. As agents are ex-ante homogeneous, the principal is indifferent over which specific subset of agents a monitor observes, conditional on μ_j. Thus, all results will be expressed in terms of the measure μ_j.

For each agent $i \in I_j$, monitor j receives a productivity signal σ_{ij}. With probability $p(\mu_j) = p_0 e^{-\lambda \mu_j}$, the signal is informative, $\sigma_{ij} = \theta_i$, and with probability $1 - p(\mu_j)$, the signal is uninformative: $\sigma_{ij} = \varnothing$. I refer to $p(\mu_j)$ as monitoring efficiency. It captures both the probability of obtaining verifiable information and the legal enforceability of that information. Two primitives summarize the monitoring technology.

The first is p_0, the baseline verifiability of a task. It captures how easy it is to detect misbehavior if a monitor focuses exclusively on a single agent. High p_0 corresponds to environments where evidence of misconduct is naturally transparent—e.g. physical theft, clear performance metrics, or verifiable outputs. Low p_0 corresponds to tasks where wrongdoing is inherently hard to spot—e.g. effort in service provision, subtle quality shading, or collusion in procurement. Intuitively, p_0 measures how "detectable" shirking or corruption is, even before span-of-control considerations enter. The second is λ, the span–efficiency parameter. It governs how quickly monitoring effectiveness decays as the span of control widens. A high λ means monitors are quickly overloaded—each additional agent sharply reduces the probability of detection—so oversight is fragile to increases in span. A low λ means monitors scale more smoothly: adding agents dilutes attention only gradually. Intuitively, λ captures the elasticity of monitoring with respect to workload, or how well monitors can multitask without missing misconduct. Together, p_0 and λ define the "quality" of the monitoring technology. p_0 tells us how sharp the monitor's eyes are when focused; λ tells us how blurry those eyes become as the monitor's span grows.

After observing the signal, monitor j sends a report $r_{ij} \in \{\theta_i, \varnothing\}$ to the principal. Since multiple monitors may report on the same agent, I denote the full report vector as $r_i = (r_{ij})_{j \in M}$.

Information is assumed to be hard, i.e., verifiable by the principal at no cost. Thus, a monitor cannot misreport the signal, but can choose to withhold information by reporting $\varnothing$ regardless of the true signal.

Monitors are compensated via contracts specifying payments of the form $\int_{I_j} S(x_i, r_i)di$. They are risk neutral with utility $V(x) = x$, so expected utility is $\mathbb{E}[S(x_i, r_i)]$. Their participation constraint is:

$$\mathbb{E}[S(x_i, r_i)] \geq \bar{V}. \tag{PCM}$$

To avoid degenerate results stemming from the law of large numbers in the continuum framework, I restrict $S(x_i, r_i)$ to be a function of individual-level outcomes rather than aggregate statistics.

Finally, both agents and monitors are subject to limited liability: their wages must be weakly non-negative in all states.

Principal The principal is the residual claimant of profits. She designs the wage schemes for both agents $W(x_i, r_i)$ and monitors $S(x_i, r_i)$, and chooses the organizational structure $\Omega = (\mu_j, \mu^2)$, where μ^2 denotes the measure of overlap between monitors' spans of control i.e., the extent of *common agency.* She is risk neutral and seeks to maximize expected profits:

$$\mathbb{E}[\Pi(W, S)] = \mathbb{E}\left[\int_0^1 (\theta_i + e_i - W(x_i, r_i) - S(x_i, r_i))\, di\right],$$

where expectations are taken over the distributions of productivity θ_i and monitoring signals σ_{ij}.

The organizational design is defined as $F = (M, \Omega)$, where M is the set of hired monitors, and $\Omega = (\mu_j, \mu^2)$ specifies their individual spans of control and overlap. Agents may fall under three monitoring regimes: *unmonitored*, *singly monitored*, or *jointly monitored.*

Given the stochastic nature of both productivity and monitoring, each agent-monitor pair can be in one of four possible states:

- *State 1*: $\theta_i = \theta_L$, $\sigma_{ij} = \theta_L$.
- *State 2*: $\theta_i = \theta_L$, $\sigma_{ij} = \varnothing$.

- *State 3*: $\theta_i = \theta_H$, $\sigma_{ij} = \varnothing$.
- *State 4*: $\theta_i = \theta_H$, $\sigma_{ij} = \theta_H$.

These states correspond to four observable outputs, $X = \{x_{\theta_L,R}, x_{\theta_L,\varnothing}, x_{\theta_H,\varnothing}, x_{\theta_H,R}\}$. For any $x_i \notin X$, the agent solves:

$$e_i \in \arg\max \left\{ W(x_i, r_i) - \frac{e_i^2}{2} \right\}.$$

Under limited liability, the principal can set $W(x_i, r_i) = 0$ for $x_i \notin X$, inducing $e_i = 0$. To ensure incentive compatibility, it suffices to guarantee that:

$$W(x_i, r_i) - \frac{e_i^2}{2} \geq 0, \quad \text{for all } x_i \in X,$$

and $W(x_i, r_i) = 0$ otherwise. These constraints imply that the participation constraint is automatically satisfied in all realizations, and can therefore be omitted in subsequent analysis.

Timing The timing of the model is as follows:

1. The principal chooses the organizational design $\Omega = (\mu_j, \mu^2)$ and offers contracts $W(x_i, r_i)$, $S(x_i, r_i)$ to agents and monitors.
2. Upon acceptance, agents learn their productivity type θ_i, and monitors receive their signals σ_{ij}.
3. Monitors submit reports r_{ij}.
4. Reports arrive before effort so that verifiable audits shape incentives; agents observe the report realization when choosing e_i.
5. Output x_i is realized; payments $W(x_i, r_i)$ and $\int_{I_j} S_j(x_i, r_i) di$ are made; the principal collects the residual profits.

First-Best Benchmark Before analyzing the role of monitors and the constraints introduced by asymmetric information, it is useful to characterize the first-best allocation. Suppose that the principal can directly observe each agent's effort level. In this case, the agent's compensation can be conditioned directly on both effort and productivity, i.e., $W(e_i, \theta_i)$.

Since there is no informational asymmetry, monitoring becomes redundant and the optimal organizational design involves no hired monitors: $M = \varnothing$.

Given the ability to observe effort, the principal fully internalizes the cost of effort and solves the following optimization problem:

$$\max_{\{W,e\}} \quad \int_0^1 \mathbb{E}\left[\theta_i + e_i - W(e_i,\theta_i)\right] di$$
$$\text{s.t.} \quad W(e_i,\theta_i) \geq \frac{e_i^2}{2}, \quad \forall i \in [0,1].$$

Because effort is observable, the principal can extract all of the agents' rents by setting compensation equal to their disutility of effort, i.e.,

$$W(e_i,\theta_i) = \frac{e_i^2}{2},$$

while choosing the effort level that maximizes net surplus. The optimal choice is $e_i = 1$ for all agents and all productivity types. Thus, in the absence of informational frictions, the principal induces the efficient level of effort and pays each agent a wage equal to their effort cost, independently of their productivity level θ_i. This result hinges on the fact that marginal returns to effort are constant across productivity types.

3 Firm design under no collusion

Before studying the problem of the organizational design under collusion it is interesting to understand how the principal designs the monitoring regime and what contracts does he offers when collusion is not possible. This section is structured as follows. First, I will characterize the contracts of the agents when any monitor is hired. Then, I will characterize the contracts of one monitor whose span of control is μ and the contract of the monitored agents. I will show that the contract under states 2 and 3 (no signal is received) correspond with the contract under no monitoring. With this, I will solve the optimal span of control when there is only one monitor in the firm. I do this process for the case where two monitors are hired and crucially show the optimal span of joint monitoring discussing its effects on profits and monitor and agent behavior.

3.1 No Monitoring

In the absence of monitoring, the principal cannot distinguish between productivity types beyond the realization of output. Since productivity can be either high or low, the contract must be designed around two observable output levels: $x_{\theta_L,\varnothing}$ and $x_{\theta_H,\varnothing}$, corresponding to low and high productivity types, respectively, when no informative signal is available.

$$x_{\theta_L,\varnothing} = e_{\theta_L,\varnothing} + \theta_L, \qquad x_{\theta_H,\varnothing} = e_{\theta_H,\varnothing} + \theta_H.$$

To induce effort from both types, the principal must design a contract that satisfies the agents' incentive compatibility constraints. First, the utility of each agent type must be non-negative, ensuring participation:

$$W(x_{\theta_L,\varnothing}, \varnothing) - \frac{e^2_{\theta_L,\varnothing}}{2} \geq 0, \tag{IC1}$$

$$W(x_{\theta_H,\varnothing}, \varnothing) - \frac{e^2_{\theta_H,\varnothing}}{2} \geq 0. \tag{IC2}$$

Second, the principal must prevent high productivity agents from mimicking the behavior of low productivity agents to reduce the cost of effort. This leads to a third incentive compatibility constraint:

$$W(x_{\theta_H,\varnothing}, \varnothing) - \frac{e^2_{\theta_H,\varnothing}}{2} \geq W(x_{\theta_L,\varnothing}, \varnothing) - \frac{(e_{\theta_L,\varnothing} - \Delta\theta)^2}{2}, \tag{IC3}$$

If (IC3) does not hold, a high productivity agent could choose $\tilde{e} = e_{\theta_L,\varnothing} - \Delta\theta$ to produce $x_{\theta_L,\varnothing}$ and receive $W(x_{\theta_L,\varnothing}, \varnothing)$. Notice that constraint (IC2) becomes redundant when both (IC1) and (IC3) are met.

Given these constraints, the principal solves the following program:

$$\begin{aligned}
\max_{\{W,e\}} \quad & \mathbb{E}\left[\int_0^1 (\theta_i + e_i - W(x_i, r_i))\, di\right] \\
\text{s.t.} \quad & W(x_{\theta_L,\varnothing}, \varnothing) - \frac{e^2_{\theta_L,\varnothing}}{2} \geq 0, && \text{(IC1)} \\
& W(x_{\theta_H,\varnothing}, \varnothing) - \frac{e^2_{\theta_H,\varnothing}}{2} \geq W(x_{\theta_L,\varnothing}, \varnothing) - \frac{(e_{\theta_L,\varnothing} - \Delta\theta)^2}{2}. && \text{(IC3)}
\end{aligned}$$

The solution to this standard principal-agent problem is summarized below.

Proposition 3.1 (No Monitoring Contract)**.** *In the absence of a monitor, the contract that solves the principal's problem is given by:*

$$e_{\theta_L,\varnothing} = 1-\Delta\theta, \quad e_{\theta_H,\varnothing} = 1, \quad W(x_{\theta_L,\varnothing},\varnothing) = \frac{(1-\Delta\theta)^2}{2}, \quad W(x_{\theta_H,\varnothing},\varnothing) = \frac{1+2\Delta\theta-3\Delta\theta^2}{2}.$$

Under asymmetric information, the principal must raise the wage for high productivity agents to incentivize effort while reducing the required effort for low productivity agents. This tradeoff discourages high-type agents from mimicking low-type behavior, since doing so would require more effort relative to the output gained. As a result, low productivity agents receive zero surplus (constraint IC1 binds), while high productivity agents earn an informational rent due to the need to deter imitation (constraint IC3 also binds). These rents can be expressed as:

$$W(x_{\theta_H,\varnothing},\varnothing) - \frac{e^2_{\theta_H,\varnothing}}{2} = \Delta\theta\left(1-\frac{3}{2}\Delta\theta\right).$$

It is easy to see that the rents that high productive agents obtain are increasing in the productivity differences with their low productive counterparts, as $\Delta\theta \leq \frac{1}{2}$. To facilitate comparisons with other monitoring regimes, I define the expected profits under the no-monitoring scenario as $\Pi_{\varnothing M}$. Using the equilibrium contract:

$$\begin{aligned}\Pi_{\varnothing M} &= \int_0^1 \theta_i di + \int_0^1 e_i di - \int_0^1 W(x_i,\varnothing)di \\ &= \frac{1}{2} + \theta_L + \frac{\Delta\theta^2}{2}. \end{aligned} \tag{1}$$

This represents the minimum profit the principal can obtain in equilibrium when offering an optimal contract without any monitoring.

Perfect Monitoring Now consider the case where the principal can hire a perfect monitor, i.e., $p(\mu) = 1$ for every μ. This implies that the monitor can perfectly observe each agent's productivity within their span of control. In the absence of collusion, the principal only needs to hire a single monitor. With proper incentives, this monitor will truthfully report the productivity of all agents.

Given this, the principal offers a constant wage to the monitor, $S(x_i, r_i) = \bar{V}$, across all states of nature. This makes the monitor indifferent between reporting and not reporting in every scenario, and by assumption, she acts in the interest of the principal. The resulting organizational structure consists of one monitor overseeing all agents: formally, $M = \{m_1\}$ and $\mu = 1$.

Under perfect monitoring, the principal learns the productivity type of each agent with certainty in exchange of a fix wage. Thus, the contract offered to agents coincides with the first-best allocation: the principal sets $e_i = 1$ and transfers just enough to cover the cost of effort, $W(e_i, \theta_i) = \frac{1}{2}$. As a result, agents receive no rents from asymmetric information. The expected profits for the firm in this case are:

$$\Pi_{PM} = \frac{\theta_H + \theta_L}{2} + \frac{1}{2} - \bar{V}.$$

Comparing this to the profits under the no-monitoring regime in equation (1), the principal will prefer to hire a monitor if and only if:

$$\bar{V} \leq \frac{\Delta\theta}{2}(1 - \Delta\theta). \tag{2}$$

Even when monitoring is perfect, the decision to hire a monitor depends on both the productivity differential between agent types and the monitor's reservation wage. Specifically, for a given $\bar{V}$, a larger productivity gap $\Delta\theta$ makes it more likely that the principal will find monitoring worthwhile. This is because greater heterogeneity in agent productivity increases the informational rents that high-productivity agents can extract under asymmetric information. By hiring a monitor, the principal can eliminate these rents and increase profits—provided the cost of doing so, $\bar{V}$, is not too high.

Notably, the benefit of monitoring is maximized when $\Delta\theta\frac{1}{2}$, as this is when the term $\Delta\theta(1 - \Delta\theta)$ reaches its maximum in equation (2). At this point, agents have both a strong incentive and a high ability to misreport their type, making the losses from asymmetric information largest. At the same time, the productivity gap is still small enough that the principal values keeping both agent types in the firm, reinforcing the value of resolving the asymmetry through monitoring.

3.2 Single Monitoring

Assume now that the principal hires a single monitor to observe a subset of agents of size μ. Since there is only one monitor, I omit any subscript on the span of control. In this setting, the principal must offer two types of contracts: one to the monitor that ensures her participation, and another to the agents that deters high-productivity agents from mimicking low-productivity ones, as in previous sections.

The firm's expected profits depend on the menu of contracts for the agents, $\mathbf{W}$, the monitor's contract $\mathbf{S}$, and the monitoring structure $\Omega = \{\mu\}$. For the subset of agents not monitored, expected profits are the same as under the no-monitoring regime. For monitored agents, the principal can condition wages on both output and the monitor's report:

$$\begin{aligned}\mathbb{E}\left[\Pi(\Omega, \mathbf{W}, \mathbf{S})\right] =& (1-\mu)\Pi_{\varnothing M} + \frac{1}{2}\mu\left(\theta_H + \theta_L + p(\mu)(e_{\theta_L,\theta_L} + e_{\theta_H,\theta_H} - W(x_{\theta_L,\theta_L}, \theta_L)\right.\\ & -W(x_{\theta_H,\theta_H}, \theta_H) - S(x_{\theta_L,\theta_L}, \theta_L) - S(x_{\theta_H,\theta_H}, \theta_H) + (1-p(\mu))(e_{\theta_L,\varnothing} + e_{\theta_H,\varnothing} \\ & \left. -W(x_{\theta_L,\varnothing}, \varnothing) - W(x_{\theta_H,\varnothing}, \varnothing) - S(x_{\theta_L,\varnothing}, \varnothing) - S(x_{\theta_H,\varnothing}, \varnothing)\right)\right) \end{aligned} \quad (3)$$

Given this setup, the optimal contract for any span of control μ solves the following problem:

$$\max_{\{\mathbf{S},\mathbf{W},\mathbf{e}\}} \mathbb{E}\left[\Pi(\mu, \mathbf{W}, \mathbf{S})\right]$$

s.t.

$$S(x_i, r_i) \geq 0, \;\; i = 1,2,3,4 \quad \text{(LLS)}$$

$$\int_{I_m} \mathbb{E}(S(x_i, r_i)) \geq \bar{V}, \quad \text{(PCM)}$$

$$W(x_i, r_i) - \frac{e_i^2}{2} \geq 0, \;\; i = 1,2,4 \quad \text{(IC1)}$$

$$W(x_{\theta_H,\varnothing}, \varnothing) - \frac{e^2_{\theta_H,\varnothing}}{2} \geq W(x_{\theta_L,\varnothing}, \varnothing) - \frac{(e_{\theta_L,\varnothing} - \Delta\theta)^2}{2}, \quad \text{(IC3)}$$

In the absence of collusion, the contracts for agents and the monitor are independent, and can therefore be optimized separately. Moreover, the monitor's contract is not necessarily unique: any contract that satisfies her participation constraint and ensures weakly higher payoffs for truthful reporting in both states will be profit-maximizing.

The following proposition characterizes one such optimal contract under imperfect monitoring and no collusion:

Proposition 3.2. *(Contracts under One Monitor) If there is one monitor covering a span μ, firm profits are maximized with the following contracts:*

$$S(x_i, r_i) = \frac{\bar{V}}{\mu}, \quad W(x_{\theta_L,\theta_L}, \theta_L) = W(x_{\theta_H,\theta_H}, \theta_H) = \frac{1}{2}, \quad e_{\theta_L,\theta_L} = e_{\theta_H,\theta_H} = 1,$$

and all other variables are as specified in Proposition 3.1.

Without collusion, all reports are truthful, and the principal only needs to guarantee that the monitor's participation constraint is satisfied. Whenever the principal receives an informative signal about an agent's productivity, she can perfectly infer effort from output. In this case, informational asymmetries are resolved, and the principal implements the first-best contract. In states where $\sigma_i = \theta_i$ (hence $r_i = \theta_i$), the agent's utility can be fully extracted. However, in uninformative states, the principal must offer the contract used in the no-monitoring regime, allowing high-productivity agents to extract rents due to asymmetric information.

Using Proposition 3.3, together with equations (1) and (3), I can compute expected profits for any $\mu \in [0, 1]$:

$$\mathbb{E}\left[\Pi(\Omega, \mathbf{W}, \mathbf{S})\right] = (1-\mu)\Pi_{\varnothing M} + \mu\frac{\theta_H + \theta_L}{2} + \frac{1}{2}\mu + \mu(1-p(\mu))\frac{\Delta\theta^2 - \Delta\theta}{2} - \bar{V}.$$

Substituting the expression for $\Pi_{\varnothing M}$ yields:

$$\mathbb{E}\left[\Pi(\mu, \mathbf{W}, \mathbf{S})\right] = \Pi_{\varnothing M} - \bar{V} + \mu p(\mu)\frac{\Delta\theta(1-\Delta\theta)}{2}. \tag{4}$$

Equation (4) describes the expected profits the principal can attain for any given span of control μ. Importantly, profits depend on μ only through the product $\mu p(\mu)$, where $p(\mu)$ is decreasing in μ. The following proposition establishes the optimal span of control in this case.

Proposition 3.3. *The optimal span of control when $|M| = 1$ and there is no collusion is:*

$$\mu_1 = \min\left\{\frac{1}{\lambda}, 1\right\}.$$

This result shows that the optimal span of control depends solely on the elasticity of the monitoring technology. Intuitively, this reflects decreasing returns to scale in monitoring. When the probability of obtaining a valid signal does not deteriorate too quickly with span of control (i.e., $\lambda \leq 1$), the optimal structure is for a single monitor to oversee the entire firm. If instead, $\lambda > 1$, a part of the workforce would optimally remain unmonitored.

Corollary 3.1. *In the absence of collusion, the principal hires a monitor if and only if:*

$$\begin{cases} \frac{p_0 e^{-1}}{\lambda} \frac{\Delta\theta(1-\Delta\theta)}{2} \geq \bar{V} & \textit{if } \lambda \geq 1, \\ p_0 e^{-\lambda} \frac{\Delta\theta(1-\Delta\theta)}{2} \geq \bar{V} & \textit{if } \lambda < 1. \end{cases}$$

Both left-hand sides are:

- *Increasing in p_0 (baseline monitoring effectiveness),*
- *Decreasing in λ (monitoring elasticity),*
- *Increasing in $\Delta\theta$ for $\Delta\theta < \frac{1}{2}$.*

The decision to hire a monitor therefore depends not only on the productivity gap and the monitor's reservation utility (as in equation (2)), but also on the efficiency of the monitoring technology. Specifically, a monitor is more likely to be hired when (i) baseline detection ability is high (p_0 is large), and (ii) monitoring efficiency declines slowly with span of control (i.e., λ is low). Consequently, tasks with low verifiability—such as creative or knowledge-intensive activities—are less likely to be monitored under this structure. In contrast, when monitoring is scalable and robust to increased span, it is optimal for a single monitor to supervise a large share of the workforce.

3.3 Double Monitoring

Assume now that the principal hires two distinct monitors, m_1 and m_2. To define the organizational structure Ω, the principal must choose three variables: the span of control for each monitor, μ_{m_1} and μ_{m_2}, and the extent of joint supervision, μ_2. Note that the form of the monitors' contracts remains unchanged. This is because the only constraint the principal faces when designing these contracts is the monitors' participation constraint. Consequently, for any given span of control μ_{m_i}, offering a constant wage equal to $\frac{\bar{V}}{\mu_{m_i}}$ for any output and report satisfies that constraint. In expectation, each monitor then receives exactly their reservation utility, independent of their span of control, making monitoring as inexpensive as possible. It follows that both monitors are optimally assigned the same span of control, i.e., $\mu_{m_1} = \mu_{m_2}$.

The next proposition shows that the optimal contracts for both the agents and the monitors do not change with the addition of a second monitor, provided collusion is not possible.

Proposition 3.4. *The optimal contract under two monitors is the same as in Proposition 3.2.*

The proof, which is omitted, follows the same logic as that of Proposition 3.2. For notational convenience, I denote the span of control of each monitor by μ_1 and the extent of joint supervision by μ^2. Since the firm's size is normalized to one, I must have $\mu_1 \in [0, 1]$. The intersection μ^2 must be non-negative and bounded such that it exceeds $2\mu_1 - 1$ and does not exceed μ_1. Thus, the feasible organizational structure must satisfy the following constraints:

$$\begin{aligned} \mu^2 &\geq 0, \\ \mu^2 &\geq 2\mu_1 - 1, \\ \mu_1 - \mu^2 &\geq 0. \end{aligned}$$

This allows me to express the expected profits under any monitoring structure $\Omega = \{\mu_1, \mu^2\}$ as:

$$\mathbb{E}(\Pi(\Omega, \mathbf{W}, \mathbf{S})) = (1 - 2\mu_1 + \mu^2)\Pi_{\varnothing M} + 2(\mu_1 - \mu^2)\Pi_{1M}(\mu_1, \mathbf{W}, \mathbf{S}) + \mu^2 \Pi_{2M}(\Omega, \mathbf{W}, \mathbf{S}),$$

where Π_{1M} and Π_{2M} represent the expected profits of the firm under single and joint monitoring, respectively. Substituting the contracts from Propositions 3.2 and 3.4 yields:

$$\mathbb{E}(\Pi(\Omega, \mathbf{W}, \mathbf{S})) = \Pi_{\varnothing M} + \left[2\mu_1 p(\mu_1) - \mu^2 p(\mu_1)^2\right] \frac{\Delta\theta(1 - \Delta\theta)}{2} - 2\bar{V}. \tag{5}$$

The profit function when two monitors are employed without collusion consists of three key components: (i) the baseline profits $\Pi_{\varnothing M}$ when no monitors are used; (ii) the cost of hiring the two monitors, $2\bar{V}$; and (iii) the informational gain, captured by the expression $[2\mu_1 p(\mu_1) - \mu^2 p(\mu_1)^2] \frac{\Delta\theta(1-\Delta\theta)}{2}$. The latter reflects the reduction in informational rents earned by high-productivity agents. Importantly, since profits are linearly decreasing in the extent of joint supervision, the following result holds:

Proposition 3.5. *The optimal length of joint supervision is given by:*

$$\mu^2 = \max\{0, 2\mu_1 - 1\}.$$

This result implies that, regardless of monitoring efficiency or span of control, the principal will always minimize the extent of joint supervision when there is no collusion. The key intuition is that assigning new agents to a monitor (increasing detection probability from 0 to $p(\mu_1)$) is more valuable than providing an additional signal on an already-monitored agent (increasing detection from $p(\mu_1)$ to $1-(1-p(\mu_1))^2$). Hence, the principal prefers to monitor more agents over duplicating signals.

The next proposition characterizes the optimal span of control as a function of λ:

Proposition 3.6.

- *If $\lambda \geq 2$, the span of control is $\mu_1 = \frac{1}{\lambda}$.*
- *If $\lambda \leq 1$, the optimal span of control is $\mu_1 = 1$.*
- *If $\lambda \in (1, 2)$:*
  - *$\mu_1 \leq \frac{1}{\lambda}$.*
  - *μ_1 is decreasing in both λ and p_0.*
  - *If $\lambda \geq 2\left(1 + W_0\left(-\frac{p_0}{e}\right)\right)$, then $\mu_1 = \frac{1}{2}$ is a local maximum of the profit function, though not necessarily a global one.*[3] *For p_0 large enough, other local maxima exist.*

The result highlights a subtle trade-off introduced by overlapping supervision. When the span of control is widened, each monitor covers more agents, and their individual detection probability declines. At the same time, however, the number of monitors observing any given agent may increase, so that detection depends not only on the effectiveness of each monitor but also on the degree of redundancy across them.

This creates what can be called an N–versus–P effect: the probability of detection is shaped both by the number of monitors (N) overseeing a task and by the effectiveness (P) of each individual monitor. As μ expands, P falls because each monitor supervises more agents, but N may rise whenever spans overlap so that two monitors jointly oversee the same agent. For some values of μ, these forces offset each other, and the overall probability of detection may rise or fall depending on which margin dominates. In other words, collusion deterrence is not monotone in span of control: redundancy can temporarily strengthen enforcement even as each individual monitor becomes less effective. This non-monotonicity is the central insight of the proposition.

[3] W_0 denotes the principal branch of the Lambert W function, defined implicitly by $W(x)e^{W(x)} = x$, with $W_0(-1/e) = -1$ and $W_0(0) = 0$.

I now compare profits under one and two monitors. The conditions under which the principal prefers two monitors are given in the following corollaries.

Corollary 3.2. *If $\lambda \geq 2$, the principal prefers to hire two monitors if and only if:*

$$\frac{p_0 e^{-1}}{\lambda} \cdot \frac{\Delta\theta(1-\Delta\theta)}{2} \geq \bar{V}.$$

Observe that Corollary 3.2 reproduces the same condition as Corollary 3.1. This is because, when the optimal span of control is less than one half, and the principal is willing to hire one monitor to supervise $\frac{1}{\lambda}$ agents for a wage $\bar{V}$, then it is also optimal to hire a second monitor to cover the remainder of the firm. More generally, this suggests that if the condition in the corollary holds and $\lambda \geq n$, it is optimal to hire n monitors. However, this does not necessarily extend to cases where $\mu_1 > \frac{1}{2}$, since the marginal benefit of hiring an additional monitor declines due to the joint supervision effect. The next corollary characterizes the hiring decision as a function of the monitors' reservation utility and monitoring efficiency when $\lambda \leq 1$.

Corollary 3.3. *If $\lambda \leq 1$, the principal hires two monitors if and only if:*

$$p_0 e^{-\lambda}\left(1 - p_0 e^{-\lambda}\right) \cdot \frac{\Delta\theta(1-\Delta\theta)}{2} \geq \bar{V}.$$

This corollary shows that, under both very low and very high monitoring efficiencies, the incentive to hire two monitors is weak. When detection is poor, no monitor is valuable enough relative to their cost. Conversely, when detection is highly effective, a single monitor suffices to obtain truthful and precise information, rendering the second monitor redundant. The incentive to hire two monitors peaks at intermediate signal accuracy—particularly when the monitoring precision is around one half. This reflects a trade-off: one monitor is not sufficiently precise alone, yet signals are still informative enough to justify their cost. These regions are illustrated in Figure 1, which maps optimal hiring choices as a function of monitor efficiency and reservation utility. In this case, hiring behavior is monotonic in the cost of monitors: for fixed p_0 and λ, the number of monitors hired decreases as $\bar{V}$ increases.

Finally, I consider the hiring decision under the intermediate case, $\lambda \in (1, 2)$.

Corollary 3.4. *Assume $\lambda \in (1, 2)$. Define*

$$A(\mu_1^*; p_0, \lambda) \equiv 2\mu_1^* p(\mu_1^*) - \mu^2 p(\mu_1^*)^2.$$

$$V_{1M}(p_0, \lambda) \equiv \frac{p_0}{\lambda} e^{-1}\Delta, \quad V_{2M}(p_0, \lambda) \equiv \left(A(\mu_1^*; p_0, \lambda) - \tfrac{p_0}{\lambda} e^{-1}\right)\Delta.$$

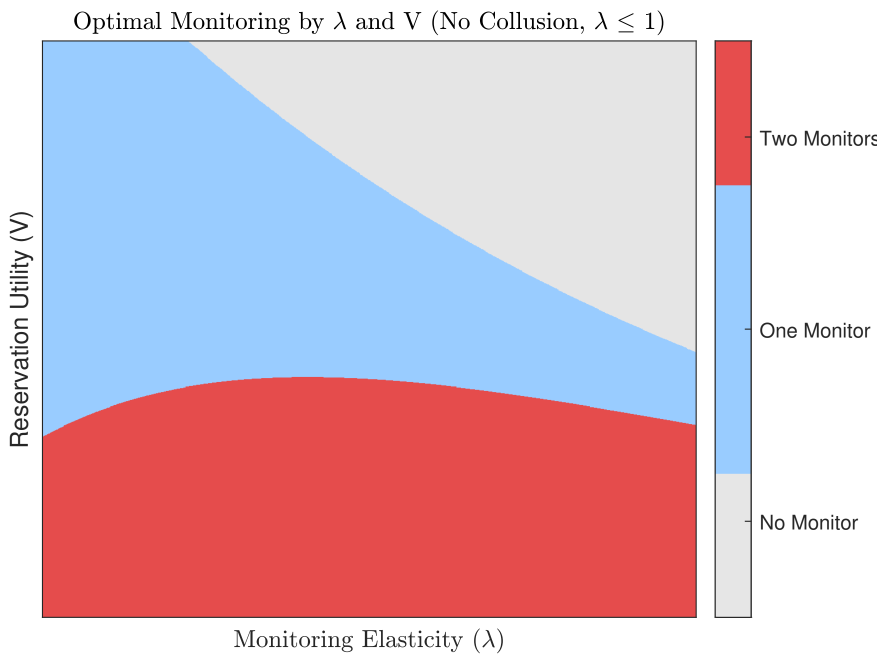

Figure 1: Optimal hiring decision for $\lambda \leq 1$ as a function of reservation wages of the monitors $\bar{V}$ and the monitoring elasticity λ. $p_0 = 0.7$. No collusion.

Thus this gives a monotonic ordering for the three hiring conditions.

$$\begin{cases} \bar{V} \leq V_2(p_0, \lambda) & \Rightarrow \textit{two monitors}, \\ V_2(p_0, \lambda) < \bar{V} < V_1(p_0, \lambda) & \Rightarrow \textit{one monitor}, \\ \bar{V} \geq V_1(p_0, \lambda) & \Rightarrow \textit{no monitor}. \end{cases}$$

Moreover, the cutoffs satisfy $\frac{\partial V_{1M}}{\partial p_0}, \frac{\partial V_{2M}}{\partial p_0} > 0$ *and* $\frac{\partial V_{1M}}{\partial \lambda}, \frac{\partial V_{2M}}{\partial \lambda} < 0$, *so higher verifiability expands the hiring regions while faster span decay contracts them.*

The hiring behavior when $\lambda \in (1, 2)$ mirrors the logic of the previous case, but with an additional trade-off. For high monitor reservation wages $\bar{V}$, monitoring is too costly and the principal hires no one. For low values of $\bar{V}$, it is optimal to employ two monitors, while for intermediate values, a single monitor is preferred.

The key difference in this range of λ is that when the span μ_1 exceeds $1/2$, some agents fall under the oversight of *two* monitors (N increases), even as each monitor's detection probability declines with span (P decreases). The net detection term,

$$A(\mu_1) = 2\mu_1 p(\mu_1) - \mu^2 p(\mu_1)^2,$$

captures precisely this tension: redundancy improves coverage but is offset by weaker individual monitors.

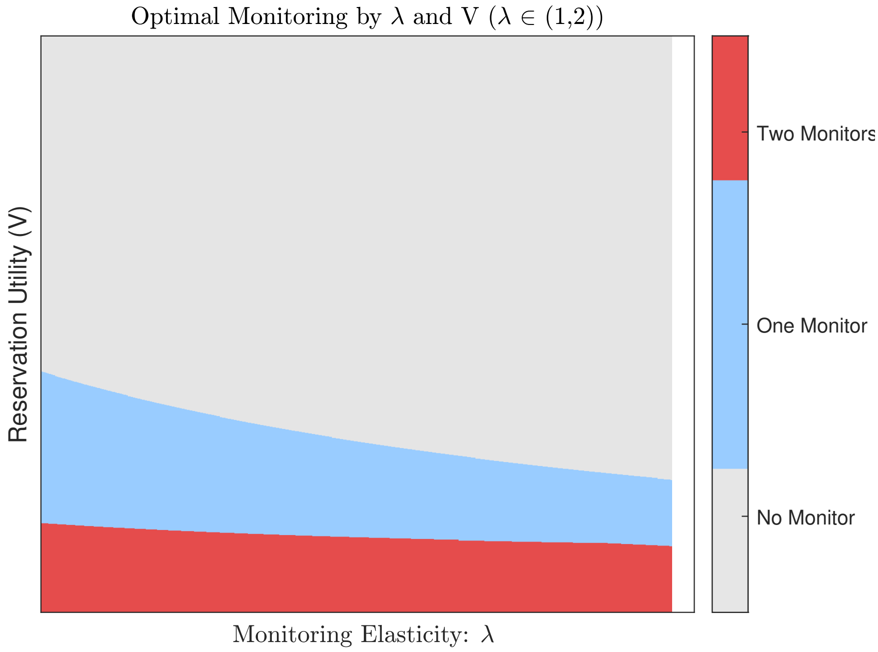

Figure 2: Optimal hiring decision for $\lambda \in (1, 2)$ as a function of reservation wages of the monitors $\bar{V}$ and the monitoring elasticity λ. $p_0 = 0.7$. No collusion.

As Figure 2 shows, comparative statics follow directly. Better verifiability ($p_0 \uparrow$) magnifies the returns to redundancy and raises the range of $\bar{V}$ for which two monitors are worthwhile. By contrast, faster decay (λ increases) erodes effectiveness more steeply as spans widen, compressing the profitability of two-monitor structures and pushing the principal toward simpler designs. Figure 2 illustrates this dynamic: unlike the $\lambda \leq 1$ case, the condition for hiring two monitors is now monotone in λ, with higher decay always making multiple monitors less attractive.

Finally, the last three corollaries help us understand how different combinations of λ and $\bar{V}$ shape the incentives to implement monitoring. When λ is relatively small, monitoring remains efficient even as the reservation wage $\bar{V}$ increases, making it optimal to hire monitors over a wide range of values for $\bar{V}$. However, as λ increases, the effectiveness of monitoring deteriorates, and the range of reservation wages for which hiring monitors is profitable narrows. In the intermediate case where $\lambda \in (1, 2)$, increases in λ not only reduce monitoring efficiency but also shrink the optimal span of control. This further limits the number of informative signals obtained, making monitoring relatively less attractive and justifiable only for low values of $\bar{V}$. The case $\lambda \geq 2$ is even more restrictive. Here, the reduction in span of control is such that part of the firm may remain entirely unmonitored. This further diminishes the returns to hiring monitors. Importantly, in this range, the conditions for hiring one or two monitors coincide and become monotonically decreasing in λ. As a result, the principal is willing to pay for monitoring only when $\bar{V}$ is sufficiently low.

In conclusion, as monitoring elasticity λ increases, the set of reservation wages $\bar{V}$ that justify the use of monitoring becomes strictly smaller. Higher values of λ therefore impose tighter constraints on the firm's ability to rely on organizational monitoring as an incentive device.

4 Firm Design under Collusion

Assume now that, after each agent learns her productivity and the monitor observes her signal, but before any effort is exerted or report is submitted to the principal, the agent and the monitor may enter into a side contract. This side contract specifies a transfer $t(x_i, r_i) \in \mathbb{R}$ from the agent to the monitor, contingent on the realized output and the report sent by the monitor. Crucially, this transfer is not restricted to be positive: a negative transfer corresponds to a payment from the monitor to the agent—for instance, to incentivize a particular effort choice.

Throughout this section, I adopt a simple assumption: regardless of the underlying bargaining process, any subset of players that can achieve a strictly higher joint utility by coordinating their actions will choose to collude.

In this context, the contracts derived in the previous section do not deter collusion. Consider the case in which a high-productivity agent is observed by a monitor who receives a perfectly informative signal. Since the monitor's wage is invariant to whether she reports or not, she is indifferent between the two actions. The agent, however, knows that if she is not reported, she will earn information rents due to asymmetric information. This creates a profitable deviation in which the agent offers a side payment to induce silence from the monitor. To eliminate this incentive, it is necessary that the combined payoff of the agent and the monitor is strictly higher when the report is truthful than when it is withheld. This requirement is central to designing contracts that are robust to collusion.

Importantly, for any contract offered by the principal, there always exists an equilibrium in which monitors truthfully report their signals—i.e., $r_{ij}(\sigma_{ij}, t(x_i, r_i)) = \sigma_{ij}$—and agents do not initiate any side transfers. This non-collusive equilibrium is sustained by the absence of credible bribing opportunities and the monitor's indifference to deviating. However, such an equilibrium may not be unique: other equilibria involving collusion can emerge, depending on the strategic incentives embedded in the contract. I define a *collusion-free contract* as a contract menu that induces a unique equilibrium in which no collusion arises, either between agents and monitors or across monitors.

When more than one monitor oversees the same agent—as in joint monitoring regimes

—collusion is not restricted to agent-monitor pairs. Under certain realizations of the signal and contract structure, one monitor may have incentives to transfer resources to another monitor in exchange for suppressing a report. In this section, I examine such possibilities and study how the threat of collusion—whether vertical (agent-monitor) or horizontal (monitor-monitor)—shapes both the optimal contract design and the internal organization of monitoring within the firm.

As in the previous section, I begin by characterizing the optimal contracts and organizational structure when only one monitor is hired. I then turn to the case of two monitors. Finally, I analyze the principal's incentives to hire monitors under varying parameter configurations, taking into account the constraints imposed by the possibility of collusion.

4.1 Single Monitoring under Collusion

Assume now that the principal hires only one monitor. As discussed earlier, the contract from Proposition 3.2 fails to be collusion-free due to incentives for high-productivity agents to collude with the monitor. In particular, if

$$S(x_{\theta_H,\theta_H}, \theta_H) \leq S(x_{\theta_H,\varnothing}, \varnothing) + W(x_{\theta_H,\varnothing}, \varnothing) - \frac{e^2_{\theta_H,\varnothing}}{2},$$

then the monitor and the agent prefer to misreport, violating collusion-freedom. This situation arises, for example, if the monitor receives the same wage regardless of whether she reports the high type or not. Thus, the contract must satisfy the following collusion constraint:

$$S(x_{\theta_H,\theta_H}, \theta_H) + W(x_{\theta_H,\theta_H}, \theta_H) - c(e_{\theta_H,\theta_H}) \geq S(x_{\theta_H,\varnothing}, \varnothing) + W(x_{\theta_H,\varnothing}, \varnothing) - c(e_{\theta_H,\varnothing}). \quad \text{(CC1)}$$

There is also a second potential collusion channel. Since the monitor's wage depends on both the report and the productivity of the agent, she might have incentives to bribe the agent to reduce effort and mimic a lower type if that increases her compensation. To prevent this, the contract must also satisfy:

$$S(x_{\theta_H,\varnothing}, \varnothing) + W(x_{\theta_H,\varnothing}, \varnothing) - c(e_{\theta_H,\varnothing}) \geq S(x_{\theta_L,\varnothing}, \varnothing) + W(x_{\theta_L,\varnothing}, \varnothing) - c(e_{\theta_L,\varnothing} - \Delta\theta). \quad \text{(CC2)}$$

Note that if the agent's incentive compatibility constraint (IC3) binds, then (CC2) reduces to

$$S(x_{\theta_H,\varnothing}, \varnothing) \geq S(x_{\theta_L,\varnothing}, \varnothing).$$

I will later verify that this is the case in equilibrium. Hence, the contract that maximizes the principal's profits under collusion must satisfy: limited liability for both agents and monitor, the monitor's participation constraint, the agents' incentive compatibility constraints (IC1 for low types and IC3 for separating types), and the collusion constraint (CC1 and CC2). The principal's problem becomes:

$$\max_{\{\mathbf{S},\mathbf{W},e\}} \ \mathbb{E}\left[\Pi(\mu,\mathbf{W},\mathbf{S})\right]$$
$$\text{s.t.} \quad \text{(LLS)}, \quad \text{(PCM)}, \quad \text{(IC1)}_i,\ i=1,2,4, \quad \text{(IC3)}, \quad \text{(CC1)}, \quad \text{(CC2)}.$$

The following proposition characterizes the optimal solution to this program.

Proposition 4.1 (Optimal Contract under One Monitor with Collusion)**.** *Suppose that agents and the monitor may collude through side contracts. Then, the profit-maximizing collusion-free contract when one monitor observes a share μ of the agents features:*

- ***Effort:*** *Agents exert first-best effort except for unreported low-productivity agents. In that case, effort is distorted downward:*

$$e_{\theta_L,\theta_L} = e_{\theta_H,\theta_H} = e_{\theta_H,\varnothing} = 1, \quad e_{\theta_L,\varnothing} = \max\left\{1 - \frac{1}{1-p(\mu)}\Delta\theta, \frac{2\bar{V}}{\Delta\theta\mu p(\mu)} + \frac{\Delta\theta}{2}\right\}.$$

- ***Agent Wages:*** *The incentive constraints (IC1) bind for low types and (IC3) for separating high types. High types reported truthfully may receive rents:*

$$W(x_{\theta_L,\theta_L},\theta_L) = \frac{1}{2}, \quad W(x_{\theta_L,\varnothing},\varnothing) = \frac{e^2_{\theta_L,\varnothing}}{2},$$
$$W(x_{\theta_H,\varnothing},\varnothing) = \frac{1}{2} + \max\left\{W(x_{\theta_L,\varnothing},\varnothing) - \frac{(e_{\theta_L,\varnothing}-\Delta\theta)^2}{2},\ S(x_{\theta_H,\theta_H},\theta_H)\right\},$$
$$W(x_{\theta_H,\theta_H},\theta_H) \geq \frac{1}{2}.$$

- ***Monitor Payments:*** *The monitor is compensated only when she truthfully reports a high-productivity agent:*

$$S(x_{\theta_H,\theta_H},\theta_H) = \max\left\{\frac{2\bar{V}}{\mu p(\mu)},\ W(x_{\theta_H,\varnothing},\varnothing) - W(x_{\theta_H,\theta_H},\theta_H)\right\}.$$

All other transfers are zero:

$$S(x_{\theta_L,\theta_L},\theta_L) = S(x_{\theta_L,\varnothing},\varnothing) = S(x_{\theta_H,\varnothing},\varnothing) = 0.$$

The key insight is that improving monitor accuracy—a higher detection probability $p(\mu)$—makes collusion more tempting, since supervisors now observe more high-productivity agents in expectation. To deter this, the principal lowers effort incentives for unreported low types, thereby shrinking the rents available for collusion. This mechanism limits bribes but comes at the cost of efficiency, since reducing effort depresses output and may lower total profits. The trade-off is thus between stronger discipline against collusion and weaker productive incentives.

We restrict attention to cases where $p(\mu)$ is small enough to ensure $e_{\theta_L,\varnothing} \geq 0$. If instead $e_{\theta_L,\varnothing} < 0$, unreported low types are excluded from the firm, and the collusion constraint simplifies: limited liability forces $S(x_{\theta_H,\theta_H}, \theta_H) = 0$, and the monitor's participation constraint binds.

To classify equilibria, let $\bar{V}_1(\mu)$ and $\bar{V}_2(\mu)$ denote the values of the monitor's reservation wage at which its participation constraint binds when (i) the collusion constraint is slack and (ii) the collusion constraint alone is binding, respectively:

$$\bar{V}_1(\mu) := \mu p(\mu)\frac{\Delta\theta}{2}\left(1 - \tfrac{3}{2}\Delta\theta\right), \qquad \bar{V}_2(\mu) := \mu p(\mu)\frac{\Delta\theta}{2}\left(1 - \Delta\theta\tfrac{3-p(\mu)}{2(1-p(\mu))}\right).$$

These thresholds partition the equilibrium space as follows:

- If $\bar{V} > \bar{V}_1(\mu)$, the monitor's participation constraint is binding while collusion is not a threat, so the non-collusion contract remains optimal.
- If $\bar{V} < \bar{V}_2(\mu)$, only the collusion constraint binds. The monitor is paid exactly enough to eliminate profitable side-payments:

$$S(x_{\theta_H,\theta_H}, \theta_H) = W(x_{\theta_H,\varnothing}, \varnothing) - W(x_{\theta_H,\theta_H}, \theta_H).$$

- If $\bar{V} \in (\bar{V}_2(\mu), \bar{V}_1(\mu))$, both constraints bind. The optimal contract then allocates fixed rents to the monitor and adjusts agent transfers accordingly:

$$W(x_{\theta_H,\theta_H}, \theta_H) = \tfrac{1}{2}, \quad S(x_{\theta_H,\theta_H}, \theta_H) = \tfrac{2\bar{V}}{\mu p(\mu)}, \quad W(x_{\theta_H,\varnothing}, \varnothing) = \tfrac{1}{2} + S(x_{\theta_H,\theta_H}, \theta_H).$$

This characterization also clarifies how rents are distributed. When $\bar{V} \leq \bar{V}_2(\mu)$, agents and monitors earn equal rents, coinciding with the agent's non-collusion rents. When $\bar{V} \in (\bar{V}_2(\mu), \bar{V}_1(\mu))$, the monitor's participation constraint absorbs all its surplus, leaving collusion rents entirely to the agent.

Importantly, monitor rents depend positively on p_0: better verifiability raises detection probability, which in turn increases the frequency of valuable reports and hence the scope for bribes. By contrast, when p_0 is low, detection is rare, collusion is less attractive, and the principal does not need to transfer as much surplus to the monitor to prevent corruption. Thus, while stronger monitoring improves accuracy, it also magnifies the rents at stake in collusion, tightening incentive constraints and altering who captures surplus in equilibrium.

4.1.1 Optimal Structure under One Monitor and Collusion

Proposition 4.1 allows us to compute the expected profits when only the collusion constraint (CC1) binds, for any given span of control:

$$\mathbb{E}\left[\Pi(\Omega, \mathbf{W}, \mathbf{S})\right] = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{4}\mu\frac{p(\mu)}{1-p(\mu)}, \tag{6}$$

and when both the collusion constraint and the monitor's participation constraint bind:

$$\Pi(\Omega, \mathbf{W}, \mathbf{S}) = \Pi_{\varnothing M} + \frac{1}{2}\mu\left[\Delta\theta(1-\Delta\theta) - (1-p(\mu))\frac{1}{2}\left(1-\frac{\Delta\theta}{2}\right)^2\right] - \frac{\bar{V}^2}{\Delta\theta^2\mu p(\mu)^2}$$
$$+ \frac{\bar{V}^2}{\Delta\theta^2\mu p(\mu)} + \frac{\bar{V}}{\Delta\theta p(\mu)} - \frac{3\bar{V}}{2p(\mu)} + \bar{V}\left(\frac{1}{2} - \frac{1}{\Delta\theta}\right).$$

Notice that if $\bar{V} = \bar{V}_2(\mu)$ both expressions coincide, which implies that, fixing μ the profit function is continuous with respect to the reservation utility of the monitor. The same reasoning applies when $\bar{V} = \bar{V}_1$. This makes the profit function continuous for $\bar{V}$. The next proposition characterizes the optimal monitoring span under collusion when $\bar{V} < \bar{V}_2$:

Proposition 4.2. *Assume that in the optimal program, PCM is slack. Then, the optimal span of control μ that maximizes profits when one monitor is hired under collusion is implicitly given by:*

$$\mu = \min\left\{\frac{1-p_0e^{-\lambda\mu}}{\lambda}, 1\right\},$$

which has the explicit solution using the main branch of the Lambert W function, W_0:

$$\mu = \min\left\{\frac{1+W_0(-p_0e^{-1})}{\lambda}, 1\right\}.$$

This proposition highlights that optimal span of control shrinks in the presence of collusion. The reason is that higher monitoring efficiency—through either larger λ or p_0—increases the exposure of the monitor to high-type agents. This, in turn, raises the incentive and the cost to prevent collusion. The principal internalizes this cost by reducing the number of agents monitored, even if monitoring is costless and reporting is frictionless. In short, under collusion, the cost of a monitor becomes increasing in their span of control.

Corollary 4.1 (Optimal Span of Control under Collusion)**.** *The optimal span of control under collusion is strictly smaller than under non-collusion, except in the special case where both coincide. In particular, the two are equal—and equal to one—if and only if:*

$$\lambda \leq 1 + W_0(-p_0 e^{-1}),$$

where $W_0(\cdot)$ denotes the main branch of the Lambert W function.

Moreover, the optimal span of control μ^* is strictly decreasing in both the elasticity of monitoring λ and the baseline detection probability p_0. In the limit as p_0 converges to 1, the optimal μ^* converges to zero.

As in the non-collusion case, a higher monitoring elasticity λ reduces the optimal span of control. However, under collusion, even p_0—the baseline precision of monitoring—has a negative effect on μ^*. A higher p_0 increases the chance of collusion occurring and hence raises the compensation needed to deter it. This prompts the principal to reduce μ until the firm re-enters the region where collusion is no longer profitable.

Hiring vs Not Hiring a Monitor under Collusion. Proposition 4.1 also allows me to compare the profitability of hiring a monitor with the outside option of not hiring at all. When only (CC1) is binding and $\mu \leq 1$, the expected profits can be written as:

$$\mathbb{E}\left[\Pi(\mu, \mathbf{W}, \mathbf{S})\right] = \Pi_{\varnothing M} - \frac{\Delta\theta^2}{4\lambda} W\left(-p_0 e^{-1}\right).$$

As in the non-collusion case, profits fall with λ and rise with p_0 and $\Delta\theta$. Higher baseline probabilities p_0 raise detection rates and thus improve monitoring effectiveness, which boosts profits via (i) an adjustment in the optimal span of control, and (ii) a direct efficiency gain. The role of λ is subtler but follows a similar logic. As $W_0\left(-p_0 e^{-1}\right) < 0$, profits are decreasing in λ; when the elasticity of monitoring decreases, profits increase.

Notice that since $W(-p_0 e^{-1}) \leq 0$, the firm always achieves *strictly higher* profits by hiring a monitor than not, provided that the monitor's cost satisfies $\bar{V} \leq \bar{V}_2$. Assume now that $\lambda \leq 1 + W_0\left(-p_0 e^{-1}\right)$. From the Corollary, we know that $\mu = 1$. Hence, the profit function can be written as,

$$\mathbb{E}\left[\Pi(1, \mathbf{W}, \mathbf{S})\right] = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{4} \frac{p_0 e^{-\lambda}}{1 - p_0 e^{-\lambda}}.$$

As the second term is positive, also in this case the principal prefers to hire one monitor than none. As a consequence, the following Corollary.

Corollary 4.2. *When $\bar{V} \leq \bar{V}_2$, the principal always prefers to hire one monitor than none.*

4.2 Double Monitoring

When two monitors are employed and agents face different monitoring regimes, the principal must design distinct contracts tailored to each situation. Specifically, agents may fall into one of three regimes: unmonitored, monitored by a single monitor, or jointly monitored. Accordingly, the principal must offer a separate contract for each case. The same applies to monitors: each is given a contract for the region they supervise independently, and a different contract for the region of joint supervision.

If the monitor's participation constraint binds, the principal pays the monitor the minimum required wage. It is more interesting, however, to analyze the case where the participation constraint is slack. In this case, the principal may condition the monitor's wage not only on her own report but also on the report of the other monitor. In the joint monitoring region, the principal can implement cross-payment schemes with penalties ($\beta \geq 0$) and bonuses ($\delta \geq 0$) to incentivize honest reporting. I will index contracts by the number of monitors involved in the supervision.

Hiring two monitors generates two main benefits. First, as in the single-monitoring case, it increases the likelihood of detecting agents' productivity. Second, through cross-monitoring and associated wage structures, it mitigates the incentives to collude. Formally, the firm's expected profit when hiring two monitors is:

$$\mathbb{E}\left[\Pi(\Omega, \mathbf{W}, \mathbf{S})\right] = (1 - 2\mu_1 + \mu^2)\Pi_{\varnothing M} + 2(\mu_1 - \mu^2)\Pi_{1M} + \mu^2\Pi_{2M},$$

where Π_{iM} denotes the profits in the region monitored by i monitors. Note that if $\mu^2 = 0$, the monitors do not overlap, and the profit function becomes a linear transformation of the one-monitor case. Hence, when there is no overlap, each monitor's span of control is identical to the single-monitor case.

Conversely, when $\mu^2 = \mu_1$, both monitors observe the same agents. Then the profit function simplifies to:

$$\mathbb{E}\left[\Pi((\mu_1, \mu_1), \mathbf{W}, \mathbf{S})\right] = (1 - \mu_1)\Pi_{\varnothing M} + \mu_1\Pi_{2M},$$

which mirrors the one-monitor structure but replaces single-monitor profits with joint-monitor profits.

Before solving the principal's problem, it is crucial to understand the new sources of collusion. In regions with single monitoring, collusion constraints remain unchanged. However, in the jointly monitored region, collusion can occur in multiple ways: between one monitor and the agent, between the two monitors, or among all three parties.

To prevent this, I introduce new collusion constraints. The constraint ensuring that both monitors prefer to report truthfully when they observe the same agent is labeled (CC3):

$$2S(x_{\theta_H,\theta_H},\theta_H) + W(x_{\theta_H,\theta_H},\theta_H) - c(e_{4,1}) \geq 2S(x_{\theta_H,\varnothing},\varnothing) + W(x_{\theta_H,\varnothing},\varnothing) - c(e_{3,1}). \quad \text{(CC3)}$$

Next, (CC4) ensures that a single monitor who observes a high-type agent in the joint region prefers to report alone:

$$S(x_{\theta_H,\theta_H},\theta_H)+\delta+S(x_{\theta_H,\varnothing},\varnothing)-\beta+W(x_{\theta_H,\theta_H},\theta_H)-c(e_{4,1}) \geq 2S(x_{\theta_H,\varnothing},\varnothing)+W(x_{\theta_H,\varnothing},\varnothing)-c(e_{3,1}). \quad \text{(CC4)}$$

Lastly, (CC5) ensures that the first monitor cannot be bribed into silence when the other monitor stays silent:

$$2S(x_{\theta_H,\theta_H},\theta_H) \geq S(x_{\theta_H,\theta_H},\theta_H) + \delta + S(x_{\theta_H,\varnothing},\varnothing) - \beta. \quad \text{(CC5)}$$

Additionally, high-type agents not reported in the joint monitoring region face an incentive constraint ensuring they do not mimic low types. This constraint is denoted as $(\text{IC3})_2$.

Let the expected utility of each monitor be the weighted sum of their expected payoffs across the regions: $\mu_1 - \mu^2$ under independent monitoring and μ^2 under joint monitoring. The participation constraint ensures this expected utility exceeds $\bar{V}$. The principal's program is then:

$$\max_{\{S,W,e\}} \mathbb{E}\left[\Pi(\Omega, \mathbf{W}, \mathbf{S})\right]$$
$$\text{s.t.} \quad \text{(LLS), (PCM), (IC1), (IC3), } (\text{IC3})_2\text{, (CC1), (CC3), (CC4), (CC5), } \beta \geq 0$$

The next proposition characterizes the optimal contract under this setting.

Proposition 4.3. *Suppose that agents and the monitors may collude through side contracts. Then, the profit-maximizing collusion-free contract when $\Omega = (\mu_1, \mu_2)$ and (PCM) is slack, features:*

- ***Effort:*** *Agents exert first-best effort except for unreported low-productivity agents. In that case, effort is distorted downward:*

$$e_{\theta_L,\theta_L,2} = e_{\theta_H,\theta_H,2} = e_{\theta_H,\varnothing,2} = 1, \quad e_{\theta_L,\varnothing,2} = 1 - \frac{\Delta\theta}{(1-p(\mu_1))^2}.$$

- ***Agent Wages:*** *The incentive constraints (IC1) bind for low types and (IC3) for separating high types. High types reported truthfully may receive rents:*

$$W_2(x_{\theta_L,\theta_L}, \theta_L) = \frac{1}{2}, \quad W_2(x_{\theta_L,\varnothing}, \varnothing) = \frac{e^2_{\theta_L,\varnothing}}{2},$$
$$W_2(x_{\theta_H,\varnothing}, \varnothing) = \frac{1}{2} + \Delta\theta - \frac{2+(1-p(\mu_1))^2}{2(1-p(\mu_1))^2}\Delta\theta^2,$$
$$W_2(x_{\theta_H,\theta_H}, \theta_H) \geq \frac{1}{2}.$$

- ***Monitor Payments:*** *The monitor is compensated only when she truthfully reports a high-productivity agent:*

$$S_2(x_{\theta_H,\theta_H}, \theta_H) = \frac{1}{2}\left(W_2(x_{\theta_H,\varnothing}, \varnothing) - W_2(x_{\theta_H,\theta_H}, \theta_H)\right),$$
$$\delta = S_2(x_{\theta_H,\theta_H}, \theta_H).$$

  All other transfers are zero:

$$S_2(x_{\theta_L,\theta_L}, \theta_L) = S_2(x_{\theta_L,\varnothing}, \varnothing) = S_2(x_{\theta_H,\varnothing}, \varnothing) = 0.$$

Because all contracts satisfying the above equality yield the same profit, we can fix $W_2(x_{\theta_H,\theta_H}, \theta_H) = \frac{1}{2}$ and let $2S_2 = W_2(x_{\theta_H,\varnothing}, \varnothing) - \frac{1}{2}$. As in the previous section, this type of contract allows us to compute the most extreme case of the monitor's wage and check when the participation constraint binds.

Given the results in the previous proposition, I can write the expected profits for general monitoring regime (μ_1, μ^2) when the participation constraint is not binding as:

$$\mathbb{E}\left[\Pi(\mu, \mathbf{W}, \mathbf{S})\right] = \Pi_{\varnothing M} + \frac{\Delta\theta^2 p(\mu_1)}{2(1-p(\mu_1))}\left[\mu_1 + \mu^2 \cdot \frac{1}{1-p(\mu_1)}\right].$$

This function reveals that expected profits increase with both μ_1 and μ^2, provided the participation constraint is not binding. When it is, profits revert to a different regime where both participation constraint of the monitor and the set of collusion constraints are simultaneously binding. The next proposition formalizes the optimal overlapping monitoring structure in this context.

Notice that, if we set $\mu^2 = 0$, the previous expression becomes the profits under collusion when only one monitor is hired (6). As each monitor takes its own part of the firm, the cross payments are not present in the optimal contract.

Proposition 4.4. *The optimal monitoring structure under collusion with two monitors features a non-trivial overlap in their monitoring responsibilities. Specifically, the optimal level of joint monitoring is given by:*

$$\mu^2 = \min\{\mu_1, \tilde{\mu}_2\},$$

where $\tilde{\mu}_2$ denotes the critical level of joint monitoring that makes the monitor's participation constraint bind, conditional on a given μ_1.

The proof is immediate from the profit function.

This result stands in stark contrast to Proposition 3.5, which establishes that, in the absence of collusion, the optimal degree of monitoring overlap should be minimized. In the benchmark non-collusive case, monitors are simply redundant in the intersection, and overlapping responsibilities only duplicate costs without generating any additional incentive or informational benefits. Therefore, the principal finds it optimal to minimize μ^2 as much as possible.

Under collusion, however, the logic reverses. Allowing a non-zero overlap in monitoring coverage provides the principal with a powerful new tool: *cross-monitoring*. When two monitors observe the same agent, the principal can design contracts that depend on both monitors' reports. This allows her to structure cross-rewards that are effective at deterring collusion. These cross-incentives lower the temptation for either monitor to accept side deals from agents, because the payoff from deviating depends on the independent action of the other monitor.

Furthermore, introducing overlap can reduce the cost of achieving collusion-proofness. When monitors work in isolation, they must be individually compensated to resist collusion, which can become prohibitively expensive, especially when detection probabilities are high. But with overlapping monitoring, *deterrence is shared*: the cost of maintaining truthful behavior is spread across two agents, reducing the required transfer per monitor. As a result, the joint region can generate a net savings in enforcement costs while maintaining or increasing the principal's expected output due to higher detection likelihood. Therefore, increasing μ^2 can be strictly profit-enhancing up to the point where the participation constraint becomes binding.

This demonstrates a central insight of the model: *collusion risk reshapes the optimal organizational structure*. When collusion is a concern, the optimal design no longer minimizes redundancy—it leverages it as an anti-collusion device.

Let's assume now that $\bar{V}$ is small enough such that only the collusion constraints are binding but not the participation constraint of the monitor. In this case, using Proposition 4.4, the profit function can be rewritten as:

$$\mathbb{E}\left(\Pi(\mu, \mathbf{W}, \mathbf{S})\right) = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{2}\mu_1 \frac{(1-(1-p(\mu_1))^2)}{(1-p(\mu_1))^2}. \tag{7}$$

The structure of this profit function is exactly as in the profit function when one monitor is hired, in equation (6) but with the incremental probability of two monitors. The first term, is the profits from no monitoring. The second term is a constant times the span of control and the odds of the probability of detection. Notice that, in this case, $(1-(1-p(\mu_1)^2))$ is the probability of getting at least one informative signal with two monitors. The next proposition finds an implicit expression for the optimal span of control of one monitor.

Proposition 4.5. *Assume that, for any $\mu^2 \in (0, \mu_1)$, (PCM) is not binding. Hence, from Proposition 4.4, $\mu_1 = \mu^2$. In that case, the span of control that maximizes profits is unique and given by*

$$\mu_1 = \min\left\{1, \frac{(1-p_0e^{-\lambda\mu_1}) + (1-p_0e^{-\lambda\mu_1})^2}{2\lambda}\right\}.$$

When the monitor participation constraint (PCM) is slack for any $\mu^2 \in (0, \mu_1)$, Proposition 4.4 implies full overlap $\mu^2 = \mu_1$. The principal's objective in this case collapses to an "enforcement return" in the single span μ_1, and the first–order condition reduces to the fixed point given by the previous proposition.

The proposition isolates a regime in which the overlap is pinned down by slack PCM, so the only remaining choice is *how large* the span should be. The fixed-point form makes comparative statics and calibration transparent, and it provides a numerically stable recipe to compute μ_1^* in structural or quantitative exercises. The next proposition gives those comparative statics.

Proposition 4.6 (Comparative statics of the optimal span)**.** *Suppose* $\mu_1^* \in (0,1)$ *is an* interior *solution to*

$$\mu_1 = \frac{(1-p(\mu_1)) + (1-p(\mu_1))^2}{2\lambda}.$$

Then μ_1^* *is continuously differentiable in* (p_0, λ) *and*

$$\frac{\partial \mu_1^*}{\partial p_0} < 0, \qquad \frac{\partial \mu_1^*}{\partial \lambda} < 0.$$

A higher p_0 (better baseline detection technology) or a higher λ (faster decay of detection with the span) raise the marginal informativeness of existing coverage relative to the marginal cost of expanding it; the principal optimally *shrinks* the span. The closed-form elasticities above are directly usable for (i) local policy counterfactuals (e.g., how much to adjust μ_1 after a tech upgrade that lifts p_0), (ii) identification (sign and magnitude restrictions on estimated λ, p_0 imply testable shifts in μ_1^*), and (iii) robustness (as $\lambda \to \infty$ or $p_0 \to 1$, the derivative tends to 0, so μ_1^* tapers smoothly to smaller spans without knife-edge jumps). Together with Proposition 4.5, this delivers a complete and operational characterization of the optimal span in the two–monitor setting when overlap is endogenously maximal.

These results together with the results in Appendices A and B allow me to illustrate numerically the incentives to hire for the principal that are shown in figure 3. The figure maps the optimal monitoring regime over the parameter space $(\lambda, \bar{V})$. The parameter λ summarizes how quickly detection decays with span of control (higher λ means a given supervisor is less effective as her span widens), while $\bar{V}$ collects the stakes of a deviation (and thus the attractiveness of bribery and the required honesty premia). The dashed boundaries show the no–collusion benchmark; the solid boundaries incorporate collusion. Allowing for collusion *shrinks* the region in which a single supervisor is optimal and *expands* the region in which assigning two independent supervisors to the same worker is optimal. In other words, once bribery is feasible, thresholds shift so that the one–monitor regime is chosen less often, and redundancy becomes optimal across a wider set of $(\lambda, \bar{V})$.

Moving rightward in the figure (larger λ), supervision becomes intrinsically weak. In this range the return to redundancy vanishes and the relevant trade–off collapses to "no monitor" versus "one monitor"; the two–monitor region disappears. For moderate λ and sufficiently high $\bar{V}$, however, the two–monitor regime dominates: overlap is selected precisely where the wage savings from keeping each supervisor honest with a partner exceed the extra headcount cost. The regime boundaries thus encode a simple rule of thumb: collusion risk shifts mass from one– to two–monitor structures, while poor monitoring technology (high λ) pushes the firm toward wider spans with at most a single supervisor per worker.

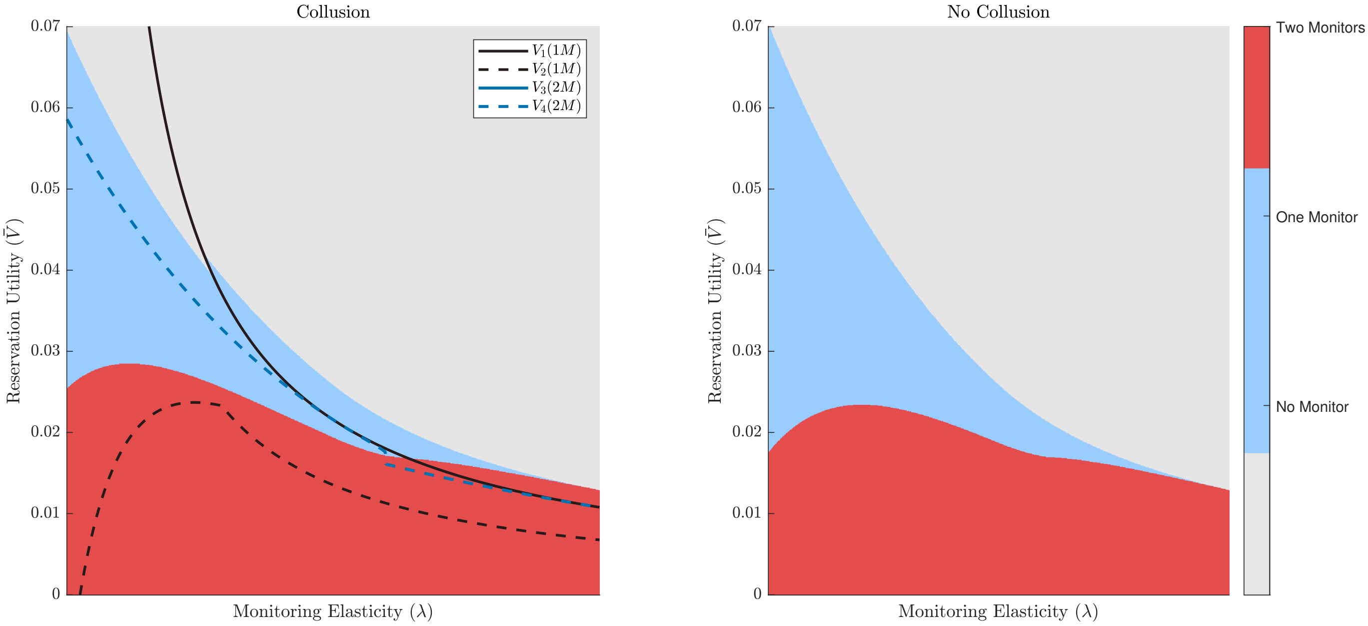

Figure 3: Regions for hiring none, one or two monitors in equilibrium. On the left, when collusion is possible. The lines V_i represent the boundaries derived. On the left, where collusion is not possible by assumption. Parameters: $\Delta\theta = 0.75$, $p_0 = 0.75$, $\lambda \in [0, 2]$.

5 Profit losses under naïve firm design

Throughout this section, assume $\bar{V}$ is low enough that only the family of collusion constraints binds. Expected profit under collusion, for a given μ_1, is

$$\mathbb{E}[\Pi(\mu, W, S)] = \Pi^{\varnothing M} + \frac{\Delta\theta^2\, p(\mu_1)}{2\,[1 - p(\mu_1)]} \left[\mu_1 + \mu^2 \cdot \frac{1}{1 - p(\mu_1)}\right], \qquad p(\mu_1) = p_0 e^{-\lambda \mu_1}.$$

As stated previously, at the optimal (full overlap, $\mu^2 = \mu_1$) this simplifies to

$$\mathbb{E}[\Pi(\mu, W, S)] = \Pi^{\varnothing M} + \frac{\Delta\theta^2}{2}\, \mu_1 \, \frac{1 - (1 - p(\mu_1))^2}{(1 - p(\mu_1))^2}.$$

Moreover, μ_1 is chosen by the fixed point

$$\mu_1 = \min\left\{ 1,\ \frac{(1 - p_0 e^{-\lambda\mu_1}) + (1 - p_0 e^{-\lambda\mu_1})^2}{2\lambda} \right\}.$$

We analyze two forms of naïveté. In the first, the principal introduces a two-monitor structure but fails to adjust contracts—she reuses the single-monitor, collusion-free contract. In the second, the principal sets the correct (collusion-free) contract but misuses organizational design, treating overlap as wasteful and minimizing it. We compare profits under both behaviors to highlight how contract design and overlap interact.

5.1 Profits under naïve contracts

When two monitors are hired, expected profit decomposes by coverage:

$$\mathbb{E}[\Pi(\Omega, \mathbf{W}, \mathbf{S})] = (1 - 2\mu_1 + \mu^2)\Pi_{\varnothing M} + 2(\mu_1 - \mu^2)\Pi_{1M} + \mu^2\Pi_{2M},$$

where $\Pi_{\varnothing M}$, Π_{1M}, and Π_{2M} denote profits in unmonitored, singly monitored, and doubly monitored regions, respectively. The principal offers the same contract everywhere—the one that solves the one-monitor problem when only the collusion constraint binds—so we compute Π_{1M} and Π_{2M} for that contract.

Singly covered region.

$$\Pi_{1M} = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{4}\,\frac{p(\mu_1)}{1 - p(\mu_1)}, \qquad p(\mu_1) = p_0 e^{-\lambda\mu_1}. \tag{8}$$

Doubly covered region. Detection succeeds with probability $1 - (1 - p(\mu_1))^2$. Reusing the one-monitor contract yields

$$\Pi_{2M} = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{4}\,\frac{p^{(2)}(\mu_1)}{1 - p^{(2)}(\mu_1)}, \qquad p^{(2)}(\mu_1) = 1 - (1 - p_0 e^{-\lambda\mu_1})^2. \tag{9}$$

Substituting (8)–(9) into the decomposition gives

$$\mathbb{E}[\Pi(\Omega, \mathbf{W}, \mathbf{S})] = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{2}\,(\mu_1 - \mu^2)\,\frac{p(\mu_1)}{1 - p(\mu_1)} + \frac{\Delta\theta^2}{4}\,\mu^2\,\frac{1 - (1 - p(\mu_1))^2}{(1 - p(\mu_1))^2}.$$

Regrouping terms,

$$\mathbb{E}[\Pi(\Omega, \mathbf{W}, \mathbf{S})] = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{2}\,\frac{p(\mu_1)}{1 - p(\mu_1)}\left(\mu_1 + \frac{p(\mu_1)}{2\,[1 - p(\mu_1)]}\,\mu^2\right).$$

Relative to the benchmark, the coefficient on μ_1 is unchanged, while the overlap term is scaled by $\frac{p(\mu_1)}{2}$. Thus, using the wrong contract attenuates (but does not reverse) the value of overlap. Since the objective is strictly increasing in μ^2, the optimal structure under naïve contracts still features full overlap ($\mu^2 = \mu_1$), and the profit simplifies to

$$\mathbb{E}[\Pi(\Omega, \mathbf{W}, \mathbf{S})] = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{4}\,\mu_1\,\frac{1 - (1 - p(\mu_1))^2}{(1 - p(\mu_1))^2}. \tag{10}$$

This coincides with the benchmark profit up to a factor of 1/2 on the monitored part, i.e., naïve contracts effectively halve the gains from monitoring.

Proposition 5.1 (Naïve-contracts optimum under single monitoring)**.** *Under naïve contracts with full overlap, any interior optimizer $\mu_1^* > 0$ satisfies*

$$\mu_1^* = \frac{\left(1 - p_0 e^{-\lambda\mu_1^*}\right)\left(2 - p_0 e^{-\lambda\mu_1^*}\right)}{2\lambda}. \tag{11}$$

Proof. With full overlap, write

$$\mathbb{E}[\Pi(\Omega, \mathbf{W}, \mathbf{S})] = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{4}\, \mu_1\, R\big(p(\mu_1)\big), \quad R(p) := \frac{p(2-p)}{(1-p)^2}, \quad p(\mu_1) = p_0 e^{-\lambda \mu_1}. \tag{12}$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}\mu_1}\Big[\mu_1 R\big(p(\mu_1)\big)\Big] = R\big(p(\mu_1)\big) - \lambda \mu_1\, p(\mu_1)\, R'\big(p(\mu_1)\big), \quad R'(p) = \frac{2}{(1-p)^3}.$$

Setting the derivative to zero and simplifying yields (11). □

For sufficiently small λ, the interior solution ceases to exist; for ease of comparison, we focus on parameter values with an interior solution. Since (10) coincides with the benchmark objective up to a constant factor, the maximizing μ_1 is the same as in the benchmark. Substituting μ_1^* into (10) gives

$$\mathbb{E}[\Pi(\Omega, \mathbf{W}, \mathbf{S})] = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{8\lambda}\, \frac{p_0 e^{-\lambda \mu_1^*}\left(2 - p_0 e^{-\lambda\mu_1^*}\right)^2}{1 - p_0 e^{-\lambda \mu_1^*}},$$

where μ_1^* is defined implicitly above.

5.2 Profits under naïve firm design

Now consider a different form of naïveté. The principal correctly implements the collusion-free contract but misperceives the role of overlap, viewing it as wasteful; hence she adopts a non-collusive monitoring structure. The resulting expected profit is

$$\mathbb{E}[\Pi(\mu, W, S)] = \Pi^{\varnothing M} + \frac{\Delta\theta^2\, p(\mu_1)}{2[1-p(\mu_1)]}\left[\mu_1 + \mu^2 \cdot \frac{1}{1-p(\mu_1)}\right], \qquad \mu^2 = \max\{0,\, 2\mu_1 - 1\}.$$

Proposition 5.2 (Naïve single-monitor design: interior optimum and value)**.** *If* $\mu^2 = 0$, *expected profits are*

$$\mathbb{E}[\Pi(\mu, W, S)] = \Pi^{\varnothing M} + \frac{\Delta\theta^2}{2}\, \frac{\mu_1\, p(\mu_1)}{1-p(\mu_1)}.$$

An interior optimizer $\mu_1^* > 0$ *(if it exists) satisfies*

$$\mu_1^* = \frac{1 - p(\mu_1^*)}{\lambda}. \tag{13}$$

Proof. Let $f(\mu_1) := \mu_1 \dfrac{p(\mu_1)}{1-p(\mu_1)}$ with $p'(\mu_1) = -\lambda p(\mu_1)$. Then

$$f'(\mu_1) = \frac{p(\mu_1)}{1-p(\mu_1)} - \mu_1\, \frac{\lambda\, p(\mu_1)}{(1-p(\mu_1))^2}.$$

Setting $f'(\mu_1^*) = 0$ yields $\mu_1^* = \dfrac{1-p(\mu_1^*)}{\lambda}$. □

Evaluating at μ_1^* gives

$$\mathbb{E}[\Pi(\mu, W, S)] = \Pi^{\varnothing M} + \frac{\Delta\theta^2}{2\lambda} \frac{p(\mu_1^*)}{1 - p(\mu_1^*)}. \tag{14}$$

For sufficiently small λ, the interior solution fails and the optimum occurs at the boundary $\mu_1^* = \frac{1}{2}$. Below we focus on parameter values for which the interior maximum exists.

Assume now $\mu^2 = 2\mu_1 - 1$ (i.e., $\mu_1 > \frac{1}{2}$). Then

$$\mathbb{E}[\Pi(\mu, W, S)] = \Pi^{\varnothing M} + \frac{\Delta\theta^2\, p(\mu_1)}{2\,[1 - p(\mu_1)]^2}\Big[3\mu_1 - 1 - \mu_1 p(\mu_1)\Big].$$

Proposition 5.3 (Optimal span μ_1^*)**.** *Any interior maximizer* $\mu_1^* \in \left(\frac{1}{2}, 1\right)$ *satisfies*

$$\mu_1^* = \frac{(1 - p(\mu_1^*))(3 - p(\mu_1^*)) + \lambda(1 + p(\mu_1^*))}{\lambda(3 + p(\mu_1^*))}, \qquad p(\mu_1^*) = p_0 e^{-\lambda \mu_1^*}.$$

If the right-hand side lies outside $\left[\frac{1}{2}, 1\right]$*, the optimum occurs at the boundary* $\mu_1^* \in \{\frac{1}{2}, 1\}$*.*

Proof. Differentiating the objective and simplifying (using $p'(\mu_1) = -\lambda p(\mu_1)$) yields the stated fixed point; details are as in the text. □

Substituting μ_1^* into the profit function,

$$\mathbb{E}[\Pi(\mu, W, S)] = \Pi^{\varnothing M} + \frac{\Delta\theta^2}{2} \frac{p_0 e^{-\lambda\mu_1^*}}{(1 - p_0 e^{-\lambda\mu_1^*})^2} \Bigg[3\, \frac{(1 - p_0 e^{-\lambda\mu_1^*})(3 - p_0 e^{-\lambda\mu_1^*}) + \lambda(1 + p_0 e^{-\lambda\mu_1^*})}{\lambda(3 + p_0 e^{-\lambda\mu_1^*})} - 1$$
$$- \frac{(1 - p_0 e^{-\lambda\mu_1^*})(3 - p_0 e^{-\lambda\mu_1^*}) + \lambda(1 + p_0 e^{-\lambda\mu_1^*})}{\lambda(3 + p_0 e^{-\lambda\mu_1^*})} p_0 e^{-\lambda\mu_1^*}\Bigg],$$

which simplifies to

$$\mathbb{E}[\Pi(\mu, W, S)] = \Pi^{\varnothing M} + \frac{\Delta\theta^2}{2\lambda} \frac{p_0 e^{-\lambda\mu_1^*}\left[\lambda p_0 e^{-\lambda\mu_1^*} + (p_0 e^{-\lambda\mu_1^*} - 3)^2\right]}{(1 - p_0 e^{-\lambda\mu_1^*})(3 + p_0 e^{-\lambda\mu_1^*})}.$$

5.2.1 Quantifying losses

Benchmark versus naïve design. Under the benchmark, the principal jointly optimizes incentives and organization. She sets full overlap ($\mu^2 = \mu_1$), maximizing enforcement depth: the probability at least one monitor detects wrongdoing is $1 - (1 - p(\mu_1))^2$, and profits increase in this term. The objective is

$$\Pi_{\varnothing M} + \frac{\Delta\theta^2}{2}\, \mu_1\, \frac{p(\mu_1)\,[2 - p(\mu_1)]}{(1 - p(\mu_1))^2},$$

and at an interior optimum μ_1^*,

$$\Pi_{\varnothing M} + \frac{\Delta\theta^2}{4\lambda} \frac{p(\mu_1^*)\,[2 - p(\mu_1^*)]^2}{1 - p(\mu_1^*)}.$$

Naïve design without overlap ($\mu^2 = 0$). If the principal forbids overlap, she allocates breadth instead of depth:

$$\Pi_{\varnothing M} + \frac{\Delta\theta^2}{2}\,\mu_1\,\frac{p(\mu_1)}{1-p(\mu_1)}, \qquad \mu_1^* = \frac{1-p(\mu_1^*)}{\lambda} \quad (\mu_1 \le \tfrac{1}{2}),$$

yielding

$$\Pi_{\varnothing M} + \frac{\Delta\theta^2}{2\lambda}\,\frac{p(\mu_1^*)}{1-p(\mu_1^*)}.$$

This discards the quadratic detection gains from redundancy; enforcement is broad but shallow, leaving each relationship vulnerable to collusion.

Allowing overlap only through $\mu^2 = 2\mu_1 - 1$ improves outcomes but remains inferior to full overlap:

$$\Pi_{\varnothing M} + \frac{\Delta\theta^2}{2}\,\frac{p(\mu_1)}{(1-p(\mu_1))^2}\,[3\mu_1 - 1 - \mu_1 p(\mu_1)],$$

with

$$\mu_1^* = \frac{(1-p(\mu_1^*))(3-p(\mu_1^*)) + \lambda(1+p(\mu_1^*))}{\lambda(3+p(\mu_1^*))}, \qquad \mathbb{E}[\Pi] = \Pi_{\varnothing M} + \frac{\Delta\theta^2}{2\lambda}\,\frac{p(\mu_1^*)\big(\lambda p(\mu_1^*) + (p(\mu_1^*)-3)^2\big)}{(1-p(\mu_1^*))(3+p(\mu_1^*))}.$$

This regime partially recovers the depth effect but cannot generally attain full overlap at the optimum.

Profits are strictly increasing in overlap: redundancy raises enforcement credibility, the scarce resource under collusion. The benchmark dominates because it tailors both incentives and architecture to maximize redundancy. Forbidding overlap ($\mu^2 = 0$) reallocates capacity to breadth, which does not cure bribability and sacrifices large detection gains. Constraining overlap ($\mu^2 = 2\mu_1 - 1$) is better than no overlap but still falls short of full overlap. The loss is organizational: misallocating monitors away from depth forfeits the non-linear increase in detection—and the associated reduction in collusion rents—that full overlap delivers.

Under both forms of naïveté, the principal suffers profit losses that stem from the same underlying mechanism: a failure to fully exploit the disciplinary power of monitoring. When contracts are naïve, monitors are not properly incentivized, so even with full overlap the firm captures only half of the potential enforcement rents. When the organization is naïve, monitoring is misallocated across workers, weakening the depth of supervision and leaving collusion more effective. In both cases, profits fall because the credibility of enforcement—the key determinant of firm discipline—is diluted, either through insufficient incentives or through an inefficient monitoring architecture.

6 Conclusion

This paper has analyzed how firms should design monitoring hierarchies when monitors themselves may collude with subordinates. Standard models treat oversight as a neutral enforcement technology; here, corruption risk makes organizational architecture part of the incentive system. The central result is that redundancy—often viewed as costly duplication—emerges as an optimal response: by overlapping monitors, principals raise the price of silence and weaken collusion.

The framework delivers three implications. First, collusion risk reshapes spans of control: narrower spans and selective overlap become optimal even when supervision technologies are otherwise efficient. Second, overlapping supervision creates new contracting instruments, since cross-monitoring contracts deter bribery at lower cost than isolated oversight. Third, redundancy helps explain why firms in high-risk environments sustain costly "four-eyes" rules or dual approvals that appear inefficient in standard models.

These insights connect to broader debates on firm organization. Work on flattening hierarchies shows that leaner structures can improve efficiency in production, but without parallel improvements in oversight they may increase vulnerability to collusion (Guadalupe and Wulf, 2010). Likewise, the model provides a new perspective on the "missing middle" in developing economies (Tybout, 2000; Macchiavello, 2012; Hsieh and Olken, 2014): if expanding scale requires costly, redundant monitoring to sustain honest enforcement, small-sized firms might find it difficult to expand.

Several avenues for research remain open. One direction is to extend the model beyond two monitors, allowing the principal to choose both the number of supervisors and the depth of monitoring layers. This would clarify the classic trade-off between flat and tall organizations: wide structures economize on oversight costs but heighten collusion risk, while deeper hierarchies provide redundancy at the expense of added layers. A second direction is to study sequential rather than simultaneous monitoring. Sequential oversight may raise detection by concentrating scrutiny on unresolved cases, but it can also introduce delays and opportunities for evidence to be concealed, creating a new speed–credibility trade-off. A third direction is empirical: linking the model to data on auditor assignment, procurement oversight, or internal controls would help quantify the costs of collusion and the returns to organizational remedies. More broadly, the analysis highlights that hierarchies are not only about efficiency or knowledge allocation: they are also instruments of enforcement, designed to remain credible even when the enforcers themselves may be tempted to collude.

References

Acemoglu, Daron and Andrew F. Newman (2002). "The Labor Market and Corporate Structure". In: *European Economic Review* 46.10, pp. 1733–1756. ISSN: 0014-2921. DOI: 10.1016/S0014-2921(01)00199-4. URL: https://doi.org/10.1016/S0014-2921(01)00199-4 (cit. on p. 3).

ACFE (2024). *Occupational Fraud 2024: Report to the Nations*. https://www.acfe.com/report-to-the-nations/2024/. Accessed: 2025-08-11. Association of Certified Fraud Examiners (cit. on p. 1).

Bandiera, Oriana et al. (2021). "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats". In: *Quarterly Journal of Economics* 136.4, pp. 2195–2242. DOI: 10.1093/qje/qjab029. URL: https://doi.org/10.1093/qje/qjab029 (cit. on p. 4).

Basel Committee on Banking Supervision (June 2011). *Principles for the Sound Management of Operational Risk*. Tech. rep. BCBS 195. Includes guidance on dual control and segregation of duties (maker–checker principle). Basel: Bank for International Settlements. URL: https://www.bis.org/publ/bcbsc102.pdf (cit. on p. 1).

Bernheim, B. Douglas and Michael D. Whinston (July 1986). "Common Agency". In: *Econometrica* 54.4, pp. 923–942. DOI: 10.2307/1912827 (cit. on p. 4).

Calvo, Guillermo A. and Stanislaw Wellisz (1978). "Supervision, Loss of Control, and the Optimum Size of the Firm". In: *Journal of Political Economy* 86.5, pp. 943–952. URL: https://www.jstor.org/stable/1828417 (cit. on p. 3).

— (1979). "Hierarchy, Ability, and Income Distribution". In: *Journal of Political Economy* 87.5, pp. 991–1010. URL: https://www.jstor.org/stable/1833180 (cit. on p. 3).

Chen, Cheng (2017). "Management Quality and Firm Hierarchy in Industry Equilibrium". In: *American Economic Journal: Microeconomics* 9.4, pp. 203–244. DOI: 10.1257/mic.20160305. URL: https://www.aeaweb.org/articles?id=10.1257/mic.20160305 (cit. on p. 3).

Chen, Cheng and Wing Suen (2019). "The Comparative Statics of Optimal Hierarchies". In: *American Economic Journal: Microeconomics* 11.2, pp. 1–25. URL: https://www.jstor.org/stable/26641414 (cit. on p. 3).

Compte, Olivier (1995). "Redundancy and Conformity in Organizations". In: *European Economic Review* 39.2, pp. 275–292. DOI: 10.1016/0014-2921(94)00052-0 (cit. on p. 4).

District of Columbia Public Schools (2023). *IMPACT: The DCPS Evaluation and Feedback System for School-Based Personnel*. Accessed: 2025-08-22. URL: https://dcps.dc.gov/page/impact-dcps-evaluation-and-feedback-system-school-based-personnel (cit. on p. 1).

Drugov, Mikhail (2010). "Competition in Bureaucracy and Corruption". In: *Journal of Development Economics* 92.2, pp. 107–114. DOI: 10.1016/j.jdeveco.2009.02.004. URL: https://doi.org/10.1016/j.jdeveco.2009.02.004 (cit. on p. 4).

Duflo, Esther et al. (2013). "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India". In: *Quarterly Journal of Economics* 128.4, pp. 1499–1545. DOI: 10.1093/qje/qjt024. URL: https://doi.org/10.1093/qje/qjt024 (cit. on p. 4).

Fan, Jingyu (2022). "Corruption Networks". In: *International Economic Review*. Revise & resubmit. URL: https://jingyufan.net/talks/ (cit. on p. 4).

Federal Deposit Insurance Corporation (2014). *Examination Policies Manual: Section 4.2 – Internal Routine and Controls*. Federal Deposit Insurance Corporation. Accessed: 2025-08-21. URL: https://www.fdic.gov/resources/supervision-and-examinations/examination-policies-manual/section4-2.pdf (cit. on p. 1).

Garicano, Luis (2000). "Hierarchies and the Organization of Knowledge in Production". In: *Journal of Political Economy* 108.5, pp. 874–904. DOI: 10.1086/317710 (cit. on pp. 1, 3).

Garicano, Luis and Esteban Rossi-Hansberg (2004). "Inequality and the Organization of Knowledge". In: *American Economic Review* 94.2, pp. 197–202. DOI: 10.1257/0002828041302037 (cit. on p. 3).

— (2006). "Organization and Inequality in a Knowledge Economy". In: *Quarterly Journal of Economics* 121.4, pp. 1383–1435. DOI: 10.1093/qje/121.4.1383. URL: https://doi.org/10.1093/qje/121.4.1383 (cit. on p. 3).

Gersen, Jacob (2007). "Overlapping and Underlapping Jurisdiction in Administrative Law". In: *University of Chicago Public Law & Legal Theory Working Papers*. Explores how overlapping authority can structure administrative incentives (cit. on p. 4).

Guadalupe, Maria and Julie Wulf (Oct. 2010). "The Flattening Firm and Product Market Competition: The Effect of Trade Liberalization on Corporate Hierarchies". In: *American Economic Journal: Applied Economics* 2.4, pp. 105–27. DOI: 10.1257/app.2.4.105. URL: https://www.aeaweb.org/articles?id=10.1257/app.2.4.105 (cit. on p. 38).

Hsieh, Chang-Tai and Benjamin A. Olken (2014). "The Missing "Missing Middle"". In: *Journal of Economic Perspectives* 28.3, pp. 89–108 (cit. on p. 38).

Ioannides, Yannis M. (2012). "Complexity and Organizational Architecture". In: *Mathematical Social Sciences* 64.2, pp. 193–202. DOI: 10.1016/j.mathsocsci.2012.04.005 (cit. on p. 3).

Keightley, Mark J. and Mark Jickling (Sept. 2017). *Banking Regulation: Overview of Regulatory Agencies*. R44918. Accessed: 2025-08-22. Congressional Research Service. URL: https://sgp.fas.org/crs/misc/R44918.pdf (cit. on p. 1).

Keren, Michael and David Levhari (1983). "The Internal Organization of the Firm and the Shape of Average Costs". In: *The Bell Journal of Economics* 14.2, pp. 474–486. URL: http://www.jstor.org/stable/3003648 (cit. on p. 3).

Kobelsky, Kevin W. (2014). "A conceptual model for segregation of duties: Integrating theory and practice for manual and IT-supported processes". In: *International Journal of Accounting Information Systems* 15.4, pp. 304–322. ISSN: 1467-0895. DOI: https:

//doi.org/10.1016/j.accinf.2014.05.003. URL: https://www.sciencedirect.com/science/article/pii/S1467089514000293 (cit. on p. 1).

Kofman, Fred and Jacques Lawarrée (1996). "A Prisoner's Dilemma Model of Collusion Deterrence". In: *Journal of Public Economics* 59.1, pp. 117–136. URL: https://doi.org/10.1016/0047-2727(95)01404-2 (cit. on p. 4).

Laffont, Jean-Jacques and David Martimort (2000). "Collusion and Delegation". In: *RAND Journal of Economics* 31.2, pp. 280–305. DOI: 10.2307/2601046 (cit. on p. 4).

Laffont, Jean-Jacques and Jean Tirole (1991). "The Politics of Government Decision-Making: A Theory of Regulatory Capture". In: *Quarterly Journal of Economics* 106.4, pp. 1089–1127. DOI: 10.2307/2937958. URL: https://doi.org/10.2307/2937958 (cit. on p. 4).

Macchiavello, Rocco (2012). "Financial Development and Vertical Integration: Theory and Evidence". In: *Journal of the European Economic Association* 10.2, pp. 255–289. DOI: 10.1111/j.1542-4774.2011.01042.x. URL: https://doi.org/10.1111/j.1542-4774.2011.01042.x (cit. on p. 38).

MacDuffie, John Paul (1995). "Human Resource Bundles and Manufacturing Performance: Organizational Logic and Flexible Production Systems in the World Auto Industry". In: *Industrial and Labor Relations Review* 48.2, pp. 197–221. DOI: 10.1177/001979399504800201. URL: https://doi.org/10.1177/001979399504800201 (cit. on p. 1).

OECD (2018). *State-Owned Enterprises and Corruption: What Are the Risks and What Can Be Done?* Paris: OECD Publishing. DOI: 10.1787/9789264303058-en. URL: https://doi.org/10.1787/9789264303058-en (cit. on p. 1).

Office of the Comptroller of the Currency (2012). *Comptroller's Handbook: Custody Services*. Tech. rep. States that "procedures should require dual control in processing of all custody assets, including securities, cash, income payments, and corporate actions." U.S. Department of the Treasury (cit. on p. 1).

Qian, Yingyi (1994). "Incentives and Loss of Control in an Optimal Hierarchy". In: *The Review of Economic Studies* 61.3, pp. 527–544. ISSN: 0034-6527. DOI: 10.2307/2297902. URL: https://doi.org/10.2307/2297902 (cit. on p. 3).

Rahman, David (2012). "But Who Will Monitor the Monitor?" In: *American Economic Review* 102.6, pp. 2767–2797. DOI: 10.1257/aer.102.6.2767. URL: https://www.aeaweb.org/articles?id=10.1257/aer.102.6.2767 (cit. on p. 4).

Reserve Bank of India (May 2011). *Implementation of Recommendations of the Committee on Customer Service in Banks*. Notification, Department of Banking Operations and Development, DBOD.No.Leg.BC.104/09.07.005/2010-11. Mandates the adoption of maker–checker (dual control) procedures in bank branches. URL: https://www.rbi.org.in/commonman/english/scripts/Notification.aspx?Id=940 (cit. on p. 1).

Rose-Ackerman, Susan (1978). *Corruption: A Study in Political Economy*. Academic Press (cit. on p. 4).

Sah, Raaj Kumar and Joseph E. Stiglitz (1986). "The Architecture of Economic Systems: Hierarchies and Polyarchies". In: *The American Economic Review* 76.4, pp. 716–727.

ISSN: 00028282. URL: http://www.jstor.org/stable/1806069 (visited on 08/07/2025) (cit. on pp. 1, 3).

Shapiro, Carl and Joseph E. Stiglitz (1984). "Equilibrium Unemployment as a Worker Discipline Device". In: *American Economic Review* 74.3, pp. 433–444. URL: https://www.jstor.org/stable/1804018 (cit. on p. 4).

Shleifer, Andrei and Robert W. Vishny (1993). "Corruption". In: *Quarterly Journal of Economics* 108.3, pp. 599–617. DOI: 10.2307/2118402 (cit. on p. 4).

Tirole, Jean (1986). "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations". In: *Journal of Law, Economics, and Organization* 2.2, pp. 181–214. DOI: 10.1093/jleo/2.2.181. URL: https://doi.org/10.1093/oxfordjournals.jleo.a036907 (cit. on pp. 1, 4, 5).

Tybout, James R. (2000). "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" In: *Journal of Economic Literature* 38.1, pp. 11–44. DOI: 10.1257/jel.38.1.11. URL: https://www.aeaweb.org/articles?id=10.1257/jel.38.1.11 (cit. on p. 38).

Williamson, Oliver E. (1967). "Hierarchical Control and Optimum Firm Size". In: *Journal of Political Economy* 75.2, pp. 123–138 (cit. on p. 3).

World Bank (2006). *Enhancing Government Effectiveness and Transparency: The Fight Against Corruption*. https://documents1.worldbank.org/curated/en/646781468330980665/pdf/Governance-and-anti-corruption-ways-to-enhance-the-World-Banks-impact.pdf. Accessed: 2025-08-11. World Bank (cit. on p. 1).

A One monitor, CC1 and PCM binding

Proposition A.1 (Shape of the optimal span in the CC + PCM regime)**.** *Fix* $\Delta\theta \in (0, \frac{1}{2})$*,* $p_0 \in (0,1)$ *and* $\bar{V} > 0$*. In the intermediate regime where the collusion constraint and the monitor participation constraint bind, the principal's profit (relative to the no–monitor baseline* $\Pi_{\varnothing M} = \frac{1}{2} + \theta_L + \frac{\Delta\theta^2}{2}$*) as a function of* $\mu \in (0,1]$ *is*

$$\begin{aligned}\Delta\theta\Pi(\mu) =& \tfrac{1}{2}\mu\left[\Delta\theta(1-\Delta\theta) - \tfrac{1}{2}\big(1-p(\mu)\big)\left(1-\tfrac{\Delta\theta}{2}\right)^2\right] \\ &+ \left(\tfrac{1}{\Delta\theta p(\mu)} - \tfrac{3}{2p(\mu)} + \tfrac{1}{2} - \tfrac{1}{\Delta}\theta\right)\bar{V} + \left(\tfrac{1}{\Delta\theta^2\mu p(\mu)} - \tfrac{1}{\Delta\theta^2\mu p(\mu)^2}\right)\bar{V}^2,\end{aligned}$$

with $p(\mu) = p_0 e^{-\lambda\mu}$*. Let* $\mu^*(\lambda) \in (0,1]$ *maximize* $\Delta\theta\Pi(\mu)$ *(assumed interior when stated). Then:*

*(i) (*Eventual collapse of span*) As* $\lambda \to \infty$*,*

$$\mu^*(\lambda) \longrightarrow 0.$$

More precisely, there exists $x^* \in (0,\infty)$ *(depending on parameters) such that* $\mu^*(\lambda) = x^*/\lambda + o(1/\lambda)$*.*

*(ii) (*Non-monotonicity in λ*) There exist parameter values and an interval of* λ *contained in the CC + PCM regime for which* $\mu^*(\lambda)$ *is non-monotonic in* λ *(it increases for some range of* λ *and decreases for larger* λ*). In particular, for* $\Delta\theta < \frac{2}{3}$ *and for moderate* $\bar{V}$ *one has* $d\mu^*/d\lambda > 0$ *at some interior optimum, while for all sufficiently large* λ *one has* $d\mu^*/d\lambda < 0$*; by continuity,* $\mu^*(\lambda)$ *must turn at least once.*

B Two monitors, all constraints binding.

We are in the two–monitor environment with spans $\mu_1 > \mu_2 \in (0,1]$. A mass $1 - 2\mu_1 + \mu_2$ is unmonitored, a mass $\mu_1 - \mu_2$ is covered by exactly one monitor, and a mass μ_2 is covered by both (overlap). Detection is $p := p(\mu_1) = p_0 e^{-\lambda\mu_1}$ and reports are independent across monitors in the overlap block. In this section, I focus on the *intermediate* region for the monitoring cost $\bar{V}$, namely $\bar{V} \in [\bar{V}_3, \bar{V}_4]$. In this region the monitor Participation Constraint (PCM) *binds* and at least one Collusion Constraint (CC) *binds*, while all low–type incentive constraints bind and the remaining constraints are slack. Intuitively, PCM pinning down the expected transfer to monitors forces the principal to use the CCs to steer the high type's behavior via truthful–report transfers, and this shows up entirely in the uninformative–state efforts $e_{2,1}$ and $e_{2,2}$ that the principal sets in the one–monitor and two–monitor blocks. The proposition below states the optimal contract in this regime and gives the closed forms of the only two endogenous efforts, $e_{2,1}$ and $e_{2,2}$, as functions of primitives and (μ_1, μ_2).

Proposition B.1 (Optimal two–monitor contract when CC & PCM bind)**.** *Fix $\mu_1 > \mu_2 \in (0,1]$ and let $p := p(\mu_1) = p_0 e^{-\lambda\mu_1}$ and $\Delta\theta \in (0, \frac{1}{2})$. Suppose $\bar{V} \in [\bar{V}_3, \bar{V}_4]$, so that the monitor Participation Constraint (PCM) and at least one Collusion Constraint is binding, while the low–type ICs bind and the other constraints are slack. Then, an optimal contract exists and is given by:*

Efforts.

$$e_{1,1} = e_{3,1} = e_{4,1} = 1, \qquad e_{1,2} = e_{3,2} = e_{4,2} = 1,$$

and the only endogenous efforts are $e_{2,1}, e_{2,2}$ determined below.

Wages to agents.

$$W^1(x_{\theta_L,\theta_L}, \theta_L) = W^2(x_{\theta_L,\theta_L}, \theta_L) = \tfrac{1}{2}, \qquad W^1(x_{\theta_L,\varnothing}, \varnothing) = \tfrac{1}{2}e_{2,1}^2, \quad W^2(x_{\theta_L,\varnothing}, \varnothing) = \tfrac{1}{2}e_{2,2}^2,$$
$$W^1(x_{\theta_H,\theta_H}, \theta_H) = W^2(x_{\theta_H,\theta_H}, \theta_H) = \tfrac{1}{2},$$
$$W^1(x_{\theta_H,\varnothing}, \varnothing) = \tfrac{1}{2} + \Delta\theta e_{2,1} - \tfrac{\Delta\theta^2}{2}, \quad W^2(x_{\theta_H,\varnothing}, \varnothing) = \tfrac{1}{2} + \Delta\theta e_{2,2} - \tfrac{\Delta\theta^2}{2}.$$

Transfers to monitors and instruments.

$$S^1(x_{\theta_H,\varnothing}, \varnothing) = S^2(x_{\theta_H,\varnothing}, \varnothing) = 0, \qquad S^1(x_{\theta_H,\theta_H}, \theta_H) = \Delta\theta\Big(e_{2,1} - \tfrac{\Delta\theta}{2}\Big),$$
$$S^2(x_{\theta_H,\theta_H}, \theta_H) = \tfrac{1}{2}\Delta\theta\Big(e_{2,2} - \tfrac{\Delta\theta}{2}\Big), \qquad \beta = 0, \qquad \delta = S^2(x_{\theta_H,\theta_H}, \theta_H).$$

Binding constraints.

- *Low–type ICs: $W^j(x_{\theta_L,\varnothing}, \varnothing) = \frac{1}{2}e_{2,j}^2$ for $j = 1, 2$.*
- *CC1, CC3, CC4, CC5 bind; with the above $(S^1, S^2, \delta, \beta)$ this is equivalent to*

$$S^1(x_{\theta_H,\theta_H}, \theta_H) = \Delta\theta e_{2,1} - \tfrac{\Delta\theta^2}{2}, \qquad 2S^2(x_{\theta_H,\theta_H}, \theta_H) = \Delta\theta e_{2,2} - \tfrac{\Delta\theta^2}{2}, \qquad \delta = S^2(x_{\theta_H,\theta_H}, \theta_H).$$

- *PCM binds:*

$$(\mu_1 - \mu_2)pS^1(x_{\theta_H,\theta_H}, \theta_H) + \mu_2 p\Big(S^2(x_{\theta_H,\theta_H}, \theta_H) + (1-p)\delta\Big) = 2\bar{V}.$$

Closed forms for $e_{2,1}$ ***and*** $e_{2,2}$***.*** *The first–order conditions together with PCM yield*

$$e_{2,1} = \frac{\dfrac{2\bar{V}}{\Delta\theta p} + (\mu_1 - \mu_2)\dfrac{\Delta\theta}{2} + \mu_2\Big(1 - \dfrac{p}{2}\Big)\Big(\dfrac{1-\Delta\theta}{1-p} + \dfrac{\Delta\theta}{2}\Big)}{(\mu_1 - \mu_2) + \mu_2\Big(1 - \dfrac{p}{2}\Big)\dfrac{2-p}{1-p}},$$

and,

$$e_{2,2} = \frac{\dfrac{2-p}{1-p}\left(\dfrac{2\bar{V}}{\Delta\theta p} + \mu_2\Big(1 - \dfrac{p}{2}\Big)\dfrac{\Delta\theta}{2}\right) + \dfrac{\mu_1 - \mu_2}{1-p}\left(\dfrac{\Delta\theta}{2}(2-p) - (1-\Delta\theta)\right)}{(\mu_1 - \mu_2) + \mu_2\Big(1 - \dfrac{p}{2}\Big)\dfrac{2-p}{1-p}}.$$

We are in the two–monitor environment with spans $\mu_1 > \mu_2 \in (0,1]$. A mass $1-2\mu_1+\mu_2$ is unmonitored, a mass $\mu_1 - \mu_2$ is covered by exactly one monitor, and a mass μ_2 is covered by both (overlap). Detection is $p := p(\mu_1) = p_0 e^{-\lambda\mu_1}$ and reports are independent across monitors in the overlap block. We focus on the *intermediate* region for the monitoring cost $\bar{V}$, namely $\bar{V} \in [\bar{V}_3, \bar{V}_4]$. In this region the monitor Participation Constraint (PCM) *binds* and at least one Collusion Constraint (CC) *binds*, while all low–type incentive constraints bind and the remaining constraints are slack. Intuitively, PCM pinning down the expected transfer to monitors forces the principal to use the CCs to steer the high type's behavior via truthful–report transfers, and this shows up entirely in the uninformative–state efforts $e_{2,1}$ and $e_{2,2}$ that the principal sets in the one–monitor and two–monitor blocks. The proposition below states the optimal contract in this regime and gives the closed forms of the only two endogenous efforts, $e_{2,1}$ and $e_{2,2}$, as functions of primitives and (μ_1, μ_2).

Let $\Delta \in (0, \frac{1}{2})$ and define

$$\Phi(e) := (1-\Delta)e + \tfrac{1}{2} - \tfrac{1}{2}e^2 + \tfrac{\Delta^2}{2}.$$

In the two-monitor intermediate regime (CC and PCM binding), with $p \equiv p(\mu_1) = p_0 e^{-\lambda\mu_1}$ and $q := 1-p$, the profit can be written as

$$\Pi(\bar{V};\mu_1,\mu_2) = (1-2\mu_1+\mu_2)\,\Pi_{\varnothing M} + (\mu_1-\mu_2)\big[p + q\,\Phi(e_{2,1})\big] + \frac{\mu_2}{2}\big[p(2-p) + q^2\,\Phi(e_{2,2})\big] - 2\bar{V}, \tag{15}$$

where $e_{2,1}$ and $e_{2,2}$ are the optimal uninformative-state efforts in the 1-monitor and 2-monitor blocks given by Proposition B.1, respectively. Note that $p = p(\mu_1)$ does not depend on μ_2. Differentiating (15) with respect to μ_2 gives

$$\frac{\partial \Pi}{\partial \mu_2} = \Pi_{\varnothing M} - \big[p + q\,\Phi(e_{2,1})\big] + \frac{1}{2}\big[p(2-p) + q^2\,\Phi(e_{2,2})\big] \\ + (\mu_1 - \mu_2)\,q\,\Phi'(e_{2,1})\,\frac{\partial e_{2,1}}{\partial \mu_2} + \frac{\mu_2}{2}\,q^2\,\Phi'(e_{2,2})\,\frac{\partial e_{2,2}}{\partial \mu_2}. \tag{16}$$

Here $\Phi'(e) = (1-\Delta) - e$.

Because $e_{2,1}$ and $e_{2,2}$ are chosen optimally given the binding constraints in this regime, the envelope theorem (for constrained problems) implies that the terms multiplying $\partial e_{2,1}/\partial \mu_2$ and $\partial e_{2,2}/\partial \mu_2$ are zero at the optimum. Therefore, at the optimal $(e^*_{2,1}, e^*_{2,2})$ we can drop the last line in (16) and write the clean envelope derivative

$$\frac{\partial \Pi}{\partial \mu_2} = \Pi_{\varnothing M} - \big[p + q\,\Phi(e^*_{2,1})\big] + \frac{1}{2}\big[p(2-p) + q^2\,\Phi(e^*_{2,2})\big]. \tag{17}$$

Sign and the role of λ. The sign of $\partial \Pi/\partial \mu_2$ in (17) depends on $p = p(\mu_1)$ and on the optimal efforts $e^*_{2,1}, e^*_{2,2}$, which themselves depend on $(\mu_1, \bar{V}, \Delta, p_0, \lambda)$ through the binding CC and PCM. Crucially, λ shapes p via $p(\mu_1) = p_0 e^{-\lambda \mu_1}$ and also affects the optimal $\mu^*_1(\lambda)$ (interior μ^*_1 scales roughly like $1/\lambda$ when the interior solution is feasible).

Intuitively:

- For small λ (slow decay), the optimal μ^*_1 is large and p is high. Then the 2-monitor term $\frac{1}{2}\big[p(2-p)+q^2\Phi(e^*_{2,2})\big]$ dominates the 1-monitor term $p+q\Phi(e^*_{2,1})$, making $\partial\Pi/\partial\mu_2 > 0$ and favoring full overlap ($\mu_2 = \mu_1$).
- For large λ (fast decay), the optimal μ^*_1 is small and p is low. The overlap advantage shrinks and the balance can flip to $\partial\Pi/\partial\mu_2 < 0$, favoring zero overlap ($\mu_2 = 0$).
- By continuity in λ and in the primitives, an interior $\mu^*_2 \in (0, \mu_1)$ solving $\partial\Pi/\partial\mu_2 = 0$ can arise for intermediate values of λ.

Hence, the sign of $\partial\Pi/\partial\mu_2$ is not globally fixed; it is governed by the detection level $p(\mu^*_1(\lambda))$ induced by λ (and by the corresponding optimal $e^*_{2,1}, e^*_{2,2}$), delivering the comparative-statics pattern you conjectured.

Regime thresholds with two monitors: $\bar{V}_3$ and $\bar{V}_4$. Now, I characterize the two thresholds $\bar{V}_3 < \bar{V}_4$ that separate these regimes.

To solve for V_4, I solve the CC-only problem and denote by (μ_1^C, μ_2^C) an optimal pair and by $p^C = p(\mu_1^C)$ the associated detection. Let E_{tot}^C be the *minimal* expected transfer to monitors that implements the CC-only optimum; PCM is slack if and only if $E_{\text{tot}}^C > \bar{V}$. Hence the boundary where PCM starts binding is

$$\bar{V}_3 = E_{\text{tot}}^C.$$

Under full overlap $\mu_1^C = \mu_2^C = \mu^C$ one obtains the closed expression

$$\bar{V}_3(\lambda, p_0, \Delta\theta) = \mu^C p^C (2 - p^C) \cdot \frac{1}{4}, \Delta\theta \left[1 - \Delta\theta, \frac{2 + (1 - p^C)^2}{2(1 - p^C)^2}\right], \tag{18}$$

where p^C is the interior CC-only optimizer (the root of the p-only FOC for that regime).

To solve for V_4, I start from the PCM-only (non-collusion) optimum and denote by (μ_1^N, μ_2^N) the optimal spans and $p^N = p(\mu_1^N)$. In the uninformative branch the second-best choice is $e_{L\varnothing} = 1 - \Delta\theta$, so the coalition CC requires a truthful transfer per informative event

$$S_{HH}^{\text{req}} = \Delta\theta e_{L\varnothing} - \tfrac{1}{2}\Delta\theta^2 = \Delta\theta - \tfrac{3}{2}\,\Delta\theta^2.$$

The cheapest way to implement this at $\mu_2^N = \max\left\{0, 2\mu_1^N - 1\right\}$, yielding a minimal total expected wage

$$\begin{aligned}\mathbb{E}(S) &= \frac{1}{2}p(\mu_1)(\mu_1 - \mu_2)S_{HH} + \frac{1}{2}p(\mu_1)\mu_2 S_{HH} \\ &= \frac{1}{2}p(\mu_1)\mu_1\Delta\theta\left(1 - \frac{3}{2}\Delta\theta\right)\end{aligned}$$

The PCM-only optimum is collusion-proof, provided $\mathbb{E}(S) = \bar{V}$, so the boundary is

$$\bar{V}_4(\lambda, p_0, \Delta\theta) = \tfrac{1}{2}\mu_1^N p_0 e^{-\lambda\mu_1^N}\left(\Delta\theta - \tfrac{3}{2}\Delta\theta^2\right). \tag{19}$$

C Proofs

Proof of Proposition 3.1.

Proof. Multiplying the objective function by two, the Lagrangian is,

$$\begin{aligned}\mathcal{L}(e_{\theta_L,\varnothing}, e_{\theta_H,\varnothing}, x_{\theta_L,\varnothing}, x_{\theta_H,\varnothing}) &= \theta_H + \theta_L + (1-q)e_{\theta_L,\varnothing} + qe_{\theta_H,\varnothing} - (1-q)W(x_{\theta_L,\varnothing}, \varnothing) \\ &- qW(x_{\theta_H,\varnothing}, \varnothing) + \lambda_1 \left(W(x_{\theta_L,\varnothing}, \varnothing) - \frac{e^2_{\theta_L,\varnothing}}{2} \right) \\ &+ \lambda_2 \left(W(x_{\theta_H,\varnothing}, \varnothing) - \frac{e^2_{\theta_H,\varnothing}}{2} - W(x_{\theta_L,\varnothing}, \varnothing) + \frac{(e_{\theta_L,\varnothing} - \Delta\theta)^2}{2} \right).\end{aligned}$$

The fist order conditions,

$$\frac{\partial \mathcal{L}}{\partial e_{\theta_L,\varnothing}} = (1-q) - \lambda_1 e_{\theta_L,\varnothing} + \lambda_2(e_{\theta_L,\varnothing} - \Delta\theta) = 0 \tag{20}$$

$$\frac{\partial \mathcal{L}}{\partial e_{\theta_H,\varnothing}} = q - \lambda_2 e_{\theta_H,\varnothing} = 0 \tag{21}$$

$$\frac{\partial \mathcal{L}}{\partial W(x_{\theta_L,\varnothing}, \varnothing)} = -1 + q + \lambda_1 - \lambda_2 = 0 \tag{22}$$

$$\frac{\partial \mathcal{L}}{\partial W(x_{\theta_H,\varnothing}, \varnothing)} = -q + \lambda_2 = 0 \tag{23}$$

From (23), $\lambda_2 = q$. Using the first order condition of $e_{\theta_H,\varnothing}$ we get that $e_{\theta_H,\varnothing} = 1$. From (22), $\lambda_1 = 1$, which implies that $W(x_{\theta_L,\varnothing}, \varnothing) = \frac{e^2_{\theta_L,\varnothing}}{2}$. At the same time, from (20),

$$(1-q) - (1-q)e_{\theta_L,\varnothing} - q\Delta\theta = 0$$

which means that

$$1 - \frac{q\Delta\theta}{1-q} = e_{\theta_L,\varnothing}$$

which is precisely $e_{\theta_H,\varnothing} - \frac{q\Delta\theta}{1-q}$. Finally, as $\lambda_2 > 0$, (IC3) is binding, which means that

$$W(x_{\theta_H,\varnothing},\varnothing) = \frac{1}{2} + \frac{e_{\theta_L,\varnothing}^2}{2} - \frac{(e_{\theta_L,\varnothing}-\Delta\theta)^2}{2} = \frac{1}{2} + \frac{\left(1-\frac{q\Delta\theta}{1-q}\right)^2}{2} - \frac{\left(\left(1-\frac{q\Delta\theta}{1-q}\right)-\Delta\theta\right)^2}{2}$$
$$= \frac{1}{2} - \frac{\Delta\theta^2 - 2\left(1-\frac{q\Delta\theta}{1-q}\right)\Delta\theta}{2}$$

□

Proof of Proposition 3.2.

Proof. It is useful to note that the problem of the monitor and the problem of the agents are not related by any inequality. Therefore, both can be solved separately. If $S(x_{\theta_L,\theta_L},\theta_L) = S(x_{\theta_L,\varnothing},\varnothing) = S(x_{\theta_H,\varnothing},\varnothing) = S(x_{\theta_H,\theta_H},\theta_H) = \frac{\bar{V}}{\mu}$ the participation constraint of the monitor holds. As in the cases where there is a report there is full information, setting $W(x_{\theta_L,\theta_L}) = W(x_{\theta_H,\theta_H}) = \frac{1}{2}$, $e_{\theta_L,\theta_L} = e_{\theta_H,\theta_H} = 1$ and the other variables as before gives the best contract.

□

Proof of Proposition 3.3

Proof. The expected profit function is

$$\mathbb{E}\left[\Pi(\mu_1,\mathbf{W},\mathbf{S})\right] = \Pi_{\varnothing M} - \bar{V} + \mu_1 p(\mu_1)\frac{\Delta\theta(1-\Delta\theta)}{2}$$

whose first order condition with respect to μ is,

$$\frac{\partial\mathbb{E}\left[\Pi(\mu,\mathbf{W},\mathbf{S})\right]}{\partial\mu} = 0 \Rightarrow p(\mu_1) + \mu_1 p'(\mu_1) = 0$$

Clearly, for $\mu_1 = 0$ the profit function is increasing, which means that the optimal span of control is always strictly positive. If there exist $\mu_1 \leq 1$ such that the first order condition is satisfied,

$$\mu_1 = -\frac{p(\mu_1)}{p'(\mu_1)}$$

else, $\mu_1 = 1$, as the first derivative of profits is always positive. A sufficient condition for μ_1 to be a global maximum is that the function $\mu_1 p(\mu_1)$ is convex everywhere, or equivalently that

$$2p'(\mu_1) + \mu_1 p''(\mu_1) < 0, \quad \mu_1 \in [0, 1]$$

In the particular case of $p(\mu_1) = p_0 e^{-\lambda \mu_1}$, the first order condition gives

$$\mu_1 = \frac{1}{\lambda}$$

and the second order condition is satisfied, as $p_0 e^{-\lambda \mu_1}$. □

Proof of Proposition 3.5

Proof. It is immediate to see from equation (5) that the profit function is linear and strictly decreasing on μ^2. Therefore μ^2 needs be larger than zero and $2\mu_1 - 1$. □

Proof of Proposition 3.6

Proof. By Proposition 3.5, if a profit-maximizing span of control satisfies $\mu_1 \leq \frac{1}{2}$, it must be a global maximum. Assume first that $\mu^2 = 0$. Then the first-order condition yields $\mu_1 = \frac{1}{\lambda}$, which is only feasible when $\lambda \geq 2$, so that $\mu_1 \leq \frac{1}{2}$.

Assume then that $\lambda < 2$. The optimal span must satisfy $\mu_1 \geq \frac{1}{2}$ and hence $\mu^2 = 2\mu_1 - 1$. The profit function becomes:

$$\Pi = \mu_1 p_0 e^{-\lambda \mu_1} - \frac{1}{2}(2\mu_1 - 1) p_0^2 e^{-2\lambda \mu_1} + \Pi_{\varnothing M} - 2\bar{V}.$$

The first-order condition is:

$$\frac{\partial \Pi}{\partial \mu_1} = p_0 e^{-\lambda \mu_1}(1 - \lambda \mu_1) - p_0^2 e^{-2\lambda \mu_1} \left(1 - \lambda(2\mu_1 - 1)\right).$$

Dividing by $p(\mu_1)$ and defining:

$$f(\mu_1; \lambda) = 1 - \lambda \mu_1 - p_0 e^{-\lambda \mu_1} \left(1 - \lambda(2\mu_1 - 1)\right),$$

we study when $f(\mu_1; \lambda) = 0$.

Assume that $\lambda = 1$. The first-order condition becomes:

$$1 - \mu_1 - p_0 e^{-\mu_1}(1 - 2\mu_1 + 1) = 0 \quad \Rightarrow \quad (1 - 2p_0 e^{-\mu_1})(1 - \mu_1) = 0.$$

This implies either $\mu_1 = 1$ or $e^{\mu_1} = 2p_0$. The first term, $(1 - 2p_0 e^{-\mu_1})$, is increasing, so if it crosses zero, the root is a local minimum. Therefore, $\mu_1 = 1$ is the global maximum.

Assume now that $\lambda \in (0, 1)$. We now show that the derivative is always positive on $\left[\frac{1}{2}, 1\right]$, so the maximum occurs at $\mu_1 = 1$. Rearranging $f(\mu_1; \lambda) \geq 0$:

$$p_0 \leq \frac{1 - \lambda\mu_1}{e^{-\lambda\mu_1}(1 - \lambda(2\mu_1 - 1))}.$$

Since $1 - \lambda(2\mu_1 - 1) \leq 1 - \lambda\mu_1$, we obtain:

$$p_0 \leq \frac{1 - \lambda\mu_1}{e^{-\lambda\mu_1}(1 - \lambda\mu_1)} = e^{\lambda\mu_1} \geq 1,$$

so $f(\mu_1; \lambda) \geq 0$ holds for all $p_0 \leq 1$, implying the maximum is at $\mu_1 = 1$.

Finally, let's study the optimal span of control when $\lambda \in (1, 2)$. First, identify when $\mu_1 = \frac{1}{2}$ is a local maximum. At $\mu_1 = \frac{1}{2}$, we get:

$$f\left(\tfrac{1}{2}; \lambda\right) = 1 - \frac{\lambda}{2} - p_0 e^{-\frac{\lambda}{2}} \leq 0 \quad \Leftrightarrow \quad p_0 \geq \left(1 - \frac{\lambda}{2}\right) e^{\frac{\lambda}{2}}.$$

Define $x = 1 - \frac{\lambda}{2} \in (0, 1)$ so the inequality becomes:

$$x e^{1-x} \leq p_0.$$

Using the Lambert W function, the condition binds when:

$$x = -W_0(-e^{-1} p_0) \quad \Rightarrow \quad \lambda = 2\left(1 + W_0\left(-\frac{p_0}{e}\right)\right).$$

Hence, if $\lambda \geq 2\left(1 + W_0\left(-\frac{p_0}{e}\right)\right)$, then $f\left(\frac{1}{2}; \lambda\right) \leq 0$, and $\mu_1 = \frac{1}{2}$ is a local maximum. In particular, if $p_0 \geq \frac{1}{2} e^{1/2} \approx 0.8244$, then this holds for any $\lambda \geq 1$. Notice that this does not rule out the possibility of having several local maximum.

In particular, I will show that when p_0 is large enough, there is at least another local maximum. If $f(1; \lambda) < 0$ and $f(\mu_1^*; \lambda) > 0$ for some $\mu_1^* \in (\frac{1}{2}, 1)$, then by the Intermediate Value Theorem there is an interior maximum. Indeed, $f(1; \lambda) = (1 - \lambda)(1 - p_0 e^{-\lambda}) < 0$ for $\lambda > 1$. At $\mu_1 = \frac{1}{2} + \frac{1}{2\lambda} \in (\frac{1}{2}, 1)$:

$$f\left(\tfrac{1}{2} + \tfrac{1}{2\lambda}; \lambda\right) = 1 - \frac{\lambda}{4} - \frac{\lambda}{2} p_0 e^{-\frac{\lambda}{2}(1 + \frac{1}{\lambda})},$$

which is positive for any $\lambda \geq 1$ and large p_0. Hence, by the Intermediate Value Theorem, there exists an interior maximum.

Now, if $\lambda < 2\left(1 + W(-p_0 e^{-1})\right)$, we can use again the Intermediate Value Theorem to show there exists an interior solution, as

$$f\left(\frac{1}{2};\lambda\right) > 0 > f\left(1;\lambda\right).$$

I will prove now that this solution is below $\frac{1}{\lambda}$. At $\mu_1 = \frac{1}{\lambda}$, we get:

$$f\left(\tfrac{1}{\lambda};\lambda\right) = -p_0 e^{-1}(\lambda - 1) < 0.$$

For $\mu_1 > \frac{1}{\lambda}$,

$$f(\mu_1;\lambda) = (1 - \lambda\mu_1)(1 - p_0 e^{-\lambda\mu_1}) - \lambda p_0 e^{-\lambda\mu_1}(1 - \mu_1) < 0,$$

as both terms are negative. So any interior solution satisfies $\mu_1 < \frac{1}{\lambda}$. Hence, if $f(\mu_1;\lambda) = 0$ it must be when $\mu_1 \in \left(\frac{1}{2}, \frac{1}{\lambda}\right)$.

To finish the proof, I will show the comparative statics using the Implicit Function Theorem. Let

$$F(\mu_1, p_0, \lambda) = 1 - \lambda\mu_1 - p_0 e^{-\lambda\mu_1}(1 - \lambda(2\mu_1 - 1)).$$

Then at the optimum $F = 0$. Its partial derivatives are, using the first order condition:

$$\begin{aligned}
\frac{\partial F}{\partial p_0} &= -\frac{1 - \lambda\mu_1}{p_0} < 0,\\
\frac{\partial F}{\partial \lambda} &= -\lambda\mu_1^2 + p_0 e^{-\lambda\mu_1}(2\mu_1 - 1),\\
\frac{\partial F}{\partial \mu_1} &= -\lambda\left[1 - 2p_0 e^{-\lambda\mu_1} + (1 - \lambda\mu_1)\right].
\end{aligned}$$

I will show that $F_\lambda < 0$. This is a decreasing function of λ. Hence, it reaches its maximum when $\lambda = 1$.

$$F_\lambda < 0 \quad \text{if and only if} \quad \frac{\mu_1^2}{2\mu_1 - 1} > p_0 e^{-\mu_1}$$

The left hand side is a decreasing function of μ_1. Hence, it reaches its minimum when $\mu_1 = 1$. As the right hand side is a probability, $1 > p_0 e^{-\mu_1}$ and the statement is always true.

Finally, let's show that $F_{\mu_1} < 0$. The condition can be rewritten as

$$2(1 - p_0 e^{-\lambda\mu_1}) > \lambda\mu_1$$

The left hand side reaches a minimum when $p_0 = 1$, and therefore, it is enough to check that

$$2 > \lambda\mu_1 + 2e^{-\lambda\mu_1}.$$

As $\mu_1 < \frac{1}{\lambda}$, define $x = \lambda\mu_1 \in \left(\frac{1}{2}, 1\right)$. The function $g(x) = x + 2e^{-x}$ has a derivative equal to

$$g'(x) = 1 - 2e^{-x}$$

which is zero at $x = \log(2)$. As the derivative has a unique solution, there is a unique critical point, that coincides with a minimum as the derivative is negative for small values of x. Hence, it is enough to compare the values of $g(x)$ at the boundaries of x.

$$g\left(\frac{1}{2}\right) = \frac{1}{2} + 2e^{-\frac{1}{2}} < g(1) = 1 + 2e^{-1} < 2$$

This proves that the partial derivative of λ is negative too. Using the Implicit Function Theorem, the partial derivatives of μ_1 are given by

$$\frac{\partial\mu_1}{\partial p_0} = -\frac{F_{p_0}}{F_{\mu_1}} = -\frac{e^{-\lambda\mu_1}(1 - \lambda(2\mu_1 - 1))}{\lambda\left[1 - 2p_0e^{-\lambda\mu_1} - p_0e^{-\lambda\mu_1}(1 - \lambda(2\mu_1 - 1))\right]}$$
$$\frac{\partial\mu_1}{\partial\lambda} = -\frac{F_\lambda}{F_{\mu_1}} = -\frac{\mu_1 + p_0e^{-\lambda\mu_1}\left[-\mu_1(1 - \lambda(2\mu_1 - 1)) - (2\mu_1 - 1)\right]}{\lambda\left[1 + p_0e^{-\lambda\mu_1}(-3 + \lambda(2\mu_1 - 1))\right]}$$

To study $\frac{\partial\mu_1}{\partial p_0}$, and $\frac{\partial\mu_1}{\partial\lambda}$ we just need to compare the signs of the two partial derivatives of F. As all three are negative, this shows that the partial derivatives are also negative when μ_1 is an interior solution. □

Proof of Corollary 3.2

Proof. Assume $\lambda \geq 2$. Then the optimal structure is given by $\Omega = \left(\frac{1}{\lambda}, 0\right)$. Comparing expected profits under this structure with $\Pi_{\varnothing M}$ yields:

$$\frac{p_0e^{-1}}{\lambda} \cdot \frac{\Delta\theta(1 - \Delta\theta)}{2} \geq \bar{V}$$

which is exactly the same condition as for hiring a single monitor under $\lambda > 2$, given in Corollary 3.1. □

Proof of Corollary 3.3

Proof. Assume now that $\lambda \leq 1$. Comparing expected profits under two-monitor and no-monitor regimes yields:

$$p_0 e^{-\lambda}\left(1 - \frac{1}{2}p_0 e^{-\lambda}\right) \cdot \frac{\Delta\theta(1-\Delta\theta)}{2} \geq \bar{V}$$

Now compare expected profits under two monitors vs. a single monitor:

$$p_0 e^{-\lambda}\left(1 - p_0 e^{-\lambda}\right) \cdot \frac{\Delta\theta(1-\Delta\theta)}{2} \geq \bar{V}$$

By comparing the two right-hand sides, we observe:

$$p_0 e^{-\lambda}\left(1 - \frac{1}{2}p_0 e^{-\lambda}\right) \geq p_0 e^{-\lambda}\left(1 - p_0 e^{-\lambda}\right)$$

Therefore, if the principal prefers two monitors over one, she also prefers two over none. However, the reverse is not necessarily true: hiring one monitor may be optimal even if two is not. This occurs when $\bar{V}$ lies between the two thresholds:

$$p_0 e^{-\lambda}\left(1 - \frac{1}{2}p_0 e^{-\lambda}\right) \cdot \frac{\Delta\theta(1-\Delta\theta)}{2} \geq \bar{V} \geq p_0 e^{-\lambda}\left(1 - p_0 e^{-\lambda}\right) \cdot \frac{\Delta\theta(1-\Delta\theta)}{2}$$

Thus, the only case in which two monitors are strictly preferred over both one and zero is when the first (larger) condition holds. □

Proof of Corollary 3.4

Proof. Assume $\lambda \in (1,2)$ and let $\Delta \equiv \frac{\Delta\theta(1-\Delta\theta)}{2}$ and write $p(\mu) = p_0 e^{-\lambda\mu}$. For $\lambda \in (1,2)$ and the optimal span $\mu_1^*(p_0, \lambda)$ in the two-monitor regime (with $\mu^2 = 2\mu_1^* - 1$), define

$$A(\mu_1^*; p_0, \lambda) \equiv 2\mu_1^* p(\mu_1^*) - \mu^2 p(\mu_1^*)^2.$$

From Corollary 3.1, the principal hires *one* monitor rather than none iff

$$\bar{V} < V_1(p_0, \lambda) \equiv \frac{p_0}{\lambda} e^{-1}\Delta.$$

Also, the principal hires *two* monitors rather than one iff

$$\bar{V} \leq V_2(p_0, \lambda) \equiv \left(A(\mu_1^*; p_0, \lambda) - \tfrac{p_0}{\lambda} e^{-1}\right)\Delta.$$

Because $A(\mu_1^*; p_0, \lambda) \geq 0$, it follows that $V_2(p_0, \lambda) \leq V_1(p_0, \lambda)$. Thus this gives a monotonic ordering for the three hiring conditions.

$$\begin{cases} \bar{V} \leq V_2(p_0, \lambda) & \Rightarrow \text{two monitors}, \\ V_2(p_0, \lambda) < \bar{V} < V_1(p_0, \lambda) & \Rightarrow \text{one monitor}, \\ \bar{V} \geq V_1(p_0, \lambda) & \Rightarrow \text{no monitor}. \end{cases}$$

Now, let's explain the comparative statics of the incentives to hire as a function fo p_0 and λ. The one-monitor cutoff is explicit:

$$\frac{\partial V_1}{\partial p_0} = \frac{e^{-1}}{\lambda}\Delta > 0, \qquad \frac{\partial V_1}{\partial \lambda} = -\frac{p_0 e^{-1}}{\lambda^2}\Delta < 0.$$

Hence better verifiability ($p_0 \uparrow$) makes hiring one monitor attractive at higher $\bar{V}$, while faster span decay ($\lambda \uparrow$) makes it less attractive.

For V_2, note that μ_1^* satisfies the first-order condition $F(\mu_1, p_0, \lambda) = 0$ from Proposition 3.5. By the Implicit Function Theorem,

$$\frac{\partial \mu_1^*}{\partial p_0} = -\frac{F_{p_0}}{F_{\mu_1}} < 0, \qquad \frac{\partial \mu_1^*}{\partial \lambda} = -\frac{F_\lambda}{F_{\mu_1}} < 0,$$

so the optimal span shrinks as either p_0 or λ rises. Differentiating V_2 shows

$$\frac{\partial V_2}{\partial p_0} > 0, \qquad \frac{\partial V_2}{\partial \lambda} < 0.$$

Thus higher p_0 expands the two-monitor region, while higher λ contracts it.

□

Proof of Proposition 4.1

Proof. First, I write the Lagrangian of the problem and its first order conditions.

Lagrangian. Let e_i, $W(x_i, r_i)$, and $S(x_i, r_i)$ denote efforts, wages to agents, and wages to the monitor respectively. The Lagrangian is:

$$
\begin{aligned}
\mathcal{L} = {} & \mathbb{E}\left[\theta_i + e_i - W(x_i, r_i) - S(x_i, r_i)\right] \\
& + \lambda_1 S(x_{\theta_L,\theta_L}, \theta_L) + \lambda_2 S(x_{\theta_L,\varnothing}, \varnothing) + \lambda_3 S(x_{\theta_H,\varnothing}, \varnothing) + \lambda_4 S(x_{\theta_H,\theta_H}, \theta_H) \\
& + \psi_1 \left(W(x_{\theta_L,\theta_L}, \theta_L) - \frac{e^2_{\theta_L,\theta_L}}{2} \right) + \psi_2 \left(W(x_{\theta_L,\varnothing}, \varnothing) - \frac{e^2_{\theta_L,\varnothing}}{2} \right) + \psi_4 \left(W(x_{\theta_H,\theta_H}, \theta_H) - \frac{e^2_{\theta_H,\theta_H}}{2} \right) \\
& + \psi_3 \left(W(x_{\theta_H,\varnothing}, \varnothing) - \frac{e^2_{\theta_H,\varnothing}}{2} - W(x_{\theta_L,\varnothing}, \varnothing) + \frac{(e_{\theta_L,\varnothing} - \Delta\theta)^2}{2} \right) \\
& + \varphi \left(p(\mu)\left(S(x_{\theta_L,\theta_L}, \theta_L) + S(x_{\theta_H,\theta_H}, \theta_H)\right) + (1 - p(\mu))\left(S(x_{\theta_L,\varnothing}, \varnothing) + S(x_{\theta_H,\varnothing}, \varnothing)\right) - \frac{2\bar{V}}{\mu} \right) \\
& + \Pi \left(S(x_{\theta_H,\theta_H}, \theta_H) + W(x_{\theta_H,\theta_H}, \theta_H) - \frac{e^2_{\theta_H,\theta_H}}{2} - S(x_{\theta_H,\varnothing}, \varnothing) - W(x_{\theta_H,\varnothing}, \varnothing) + \frac{e^2_{\theta_H,\varnothing}}{2} \right).
\end{aligned}
$$

First-order Conditions.

$$\frac{\partial \mathcal{L}}{\partial e_{\theta_L,\theta_L}} = 0 \Rightarrow \frac{p(\mu)}{2} - \psi_1 e_{\theta_L,\theta_L} = 0 \tag{24}$$

$$\frac{\partial \mathcal{L}}{\partial e_{\theta_L,\varnothing}} = 0 \Rightarrow \frac{1}{2}(1 - p(\mu)) - \psi_2 e_{\theta_L,\varnothing} + \psi_3 (e_{\theta_L,\varnothing} - \Delta\theta) = 0 \tag{25}$$

$$\frac{\partial \mathcal{L}}{\partial e_{\theta_H,\varnothing}} = 0 \Rightarrow \frac{1}{2}(1 - p(\mu)) - \psi_3 e_{\theta_H,\varnothing} + \Pi e_{\theta_H,\varnothing} = 0 \tag{26}$$

$$\frac{\partial \mathcal{L}}{\partial e_{\theta_H,\theta_H}} = 0 \Rightarrow \frac{p(\mu)}{2} - \psi_4 e_{\theta_H,\theta_H} - \Pi e_{\theta_H,\theta_H} = 0 \tag{27}$$

$$\frac{\partial \mathcal{L}}{\partial W(x_{\theta_L,\theta_L}, \theta_L)} = 0 \Rightarrow -\frac{p(\mu)}{2} + \psi_1 = 0 \tag{28}$$

$$\frac{\partial \mathcal{L}}{\partial W(x_{\theta_L,\varnothing}, \varnothing)} = 0 \Rightarrow -\frac{1}{2}(1 - p(\mu)) + \psi_2 - \psi_3 = 0 \tag{29}$$

$$\frac{\partial \mathcal{L}}{\partial W(x_{\theta_H,\varnothing}, \varnothing)} = 0 \Rightarrow -\frac{1}{2}(1 - p(\mu)) + \psi_3 - \Pi = 0 \tag{30}$$

$$\frac{\partial \mathcal{L}}{\partial W(x_{\theta_H,\theta_H}, \theta_H)} = 0 \Rightarrow -\frac{p(\mu)}{2} + \psi_4 + \Pi = 0 \tag{31}$$

$$\frac{\partial \mathcal{L}}{\partial S(x_{\theta_L,\theta_L}, \theta_L)} = 0 \Rightarrow -\frac{p(\mu)}{2} + \lambda_1 + \varphi p(\mu) = 0 \tag{32}$$

$$\frac{\partial \mathcal{L}}{\partial S(x_{\theta_L,\varnothing}, \varnothing)} = 0 \Rightarrow -\frac{1}{2}(1 - p(\mu)) + \lambda_2 + \varphi(1 - p(\mu)) = 0 \tag{33}$$

$$\frac{\partial \mathcal{L}}{\partial S(x_{\theta_H,\varnothing}, \varnothing)} = 0 \Rightarrow -\frac{1}{2}(1 - p(\mu)) + \lambda_3 - \Pi + \varphi(1 - p(\mu)) = 0 \tag{34}$$

$$\frac{\partial \mathcal{L}}{\partial S(x_{\theta_H,\theta_H}, \theta_H)} = 0 \Rightarrow -\frac{p(\mu)}{2} + \lambda_4 + \Pi + \varphi p(\mu) = 0 \tag{35}$$

From the first order condition of the efforts and the wages, it is immediate that:

$$e_{\theta_L,\theta_L} = e_{\theta_H,\theta_H} = e_{\theta_H,\varnothing} = 1.$$

From equation (28), $\psi_1 > 0$, and from (29) $\psi_2 > 0$ this implies that the constraints for low-productivity types (IC1) are binding. Hence,

$$W(x_{\theta_L,\theta_L}, \theta_L) = \frac{1}{2}, \quad W(x_{\theta_L,\varnothing}, \varnothing) = \frac{e_{\theta_L,\varnothing}^2}{2}.$$

To show that (IC3) binds, it is enough to notice that $\psi_3 > 0$ in equation (30). Then,

$$W(x_{\theta_H,\varnothing}, \varnothing) = \frac{1}{2} + W(x_{\theta_L,\varnothing}, \varnothing) - \frac{(e_{\theta_L,\varnothing} - \Delta\theta)^2}{2}.$$

And, substituting the value of $W(x_{\theta_L,\varnothing}, \varnothing)$,

$$W(x_{\theta_H,\varnothing}, \varnothing) = \frac{1}{2} + \Delta\theta e_{\theta_L,\varnothing} - \frac{\Delta\theta^2}{2}.$$

We already know that if the Participation Constraint of the Monitor is binding but the Collusion Constraint is not, we are under the non-collusion contract. Now consider the wage contract W as in Proposition 3.3, and let the monitor's wage for the truthful report be

$$S(x_{\theta_H,\theta_H}, \theta_H) = \frac{2\bar{V}}{\mu p(\mu)}.$$

This contract is incentive compatible for both agents and monitor and satisfies (CC2). If it also satisfies the first collusion constraint (CC1),

$$\bar{V} \geq \frac{1}{\lambda} p_0 e^{-1} \cdot \frac{\Delta\theta}{2} \left(1 - \frac{3}{2}\Delta\theta\right),$$

then the contract is collusion-proof. This means that the contract developed in Section 3 is also collusion proof.

Assume now that this inequality does not hold:

$$\bar{V} < \frac{1}{\lambda} p_0 e^{-1} \cdot \frac{\Delta\theta}{2} \left(1 - \frac{3}{2}\Delta\theta\right) := \bar{V}_1,$$

I explore the optimal contract in this case. Begin by assuming that the monitor's participation constraint is not binding ($\varphi = 0$); we will later examine when this assumption holds and its implications. From equations (33) and (34), $S(x_{\theta_L,\varnothing}, \varnothing) = S(x_{\theta_H,\varnothing}, \varnothing) = 0$. Hence, (CC2) reduces to (IC3) as stated before. Then, the Collusion Constraint (CC1) must bind. This implies:

$$W(x_{\theta_H,\theta_H}, \theta_H) + S(x_{\theta_H,\theta_H}, \theta_H) = W(x_{\theta_H,\varnothing}, \varnothing).$$

Solving for $e_{\theta_L,\varnothing}$:

Recall from earlier that:

$$\psi_2 = 1 - \frac{p(\mu)}{2}, \quad \Pi = \frac{p(\mu)}{2}, \quad \psi_3 + \pi_2 = \Pi + \frac{1}{2}(1 - p(\mu)).$$

The FOC for $e_{\theta_L,\varnothing}$ becomes:

$$\frac{1}{2}(1 - p(\mu)) - \psi_2 e_{\theta_L,\varnothing} + (\psi_3 + \pi_2)(e_{\theta_L,\varnothing} - \Delta\theta) = 0.$$

Substitute in and simplify:

$$\frac{1}{2}(1 - p(\mu)) - \left(1 - \frac{p(\mu)}{2}\right) e_{\theta_L,\varnothing} + \left(\frac{p(\mu)}{2} + \frac{1}{2}(1 - p(\mu))\right)(e_{\theta_L,\varnothing} - \Delta\theta) = 0.$$

Solving yields:

$$e_{\theta_L,\varnothing} = 1 - \frac{\Delta\theta}{1 - p(\mu)}.$$

This completes the derivation of the optimal contract when only the Collusion Constraint (CC1) is binding. Substituting the values into the participation constraint of the monitor, we require to have

$$\bar{V} \leq \mu p(\mu) \frac{\Delta\theta}{2}\left(1 - \Delta\theta \frac{3 - p(\mu)}{2(1 - p(\mu))}\right) := \bar{V}_2(\mu)$$

Now, suppose instead that the participation constraint of the monitor is also binding. If either $\lambda_1 > 0$ or $\lambda_2 > 0$ we can show that λ_3, $\lambda_4 > 0$ and the optimal contract is the non-collusion one. But, by assumption, we know that, in that case, (CC1) is slack. Hence, $S(x_{\theta_L,\theta_L}, \theta_L) = S(x_{\theta_L,\varnothing}, \varnothing) = 0$. Adding equations (33) and (34), we get that

$$\lambda_3 = \lambda_2 + \Pi > 0,$$

and, hence, $S(x_{\theta_H,\varnothing}, \varnothing) = 0$. Then, we can obtain the contract from the monitor by substituting $S(x_{\theta_L,\varnothing}, \varnothing) = S(x_{\theta_H,\varnothing}, \varnothing) = 0$:

$$S(x_{\theta_H,\theta_H}, \theta_H) = \frac{2\bar{V}}{\mu p(\mu)}.$$

At the same time, as (CC1) and (IC3) are binding,

$$W(x_{\theta_H,\theta_H}, \theta_H) + S(x_{\theta_H,\theta_H}, \theta_H) = \frac{1}{2} + \Delta\theta e_{\theta_L,\varnothing} - \frac{\Delta\theta^2}{2}.$$

I will distinguish two cases:

Case A. $W(x_{\theta_H,\theta_H}, \theta_H) > \frac{1}{2}$ This immediately implies that $\varphi_4 = 0$. Using a similar reasoning to before, we get $e_{\theta_L,\varnothing} = 1 - \frac{\Delta\theta}{1-p(\mu)}$. Finally, from the (IC1),

$$W(x_{\theta_H,\theta_H}, \theta_H) = \frac{1}{2} + \Delta\theta - \Delta\theta^2 \frac{3 - p(\mu)}{2(1 - p(\mu))} - \frac{2\bar{V}}{\mu p(\mu)}$$

Notice that we require that $W(x_{\theta_H,\theta_H}, \theta_H) > \frac{1}{2}$. Or equivalently,

$$\Delta\theta - \Delta\theta^2 \frac{3 - p(\mu)}{2(1 - p(\mu))} - \frac{2\bar{V}}{\mu p(\mu)} > 0$$

Which is a contradiction with the fact that $\bar{V} \geq \bar{V}_2$.

Case B. $W(x_{\theta_H,\theta_H}, \theta_H) = \frac{1}{2}$ So, from the previous two equations,

$$S(x_{\theta_H,\theta_H}, \theta_H) = \Delta\theta e_{\theta_L,\varnothing} - \frac{\Delta\theta^2}{2},$$

and

$$S(x_{\theta_L,\theta_L},\theta_L) = \frac{2\bar{V}}{\mu p(\mu)} - \Delta\theta e_{\theta_L,\varnothing} + \frac{\Delta\theta^2}{2}.$$

As $S(x_{\theta_L,\theta_L},\theta_L) = 0$,

$$e_2 = \frac{2\bar{V}}{\Delta\theta\mu p(\mu)} + \frac{\Delta\theta}{2}.$$

□

Proof of Proposition 4.2

Proof. The first order condition of the Lagrangian of the proof of Proposition 4.1 when μ is an interior solution is given by,

$$\begin{aligned}\frac{1}{2}p(\mu)((e_{\theta_L,\theta_L} + e_{\theta_H,\theta_H} - W(x_{\theta_L,\theta_L},\theta_L) - W(x_{\theta_H,\theta_H},\theta_H) - S(x_{\theta_H,\theta_H},\theta_H)\\ -(e_{\theta_H,\varnothing} + e_{\theta_L,\varnothing} - W(x_{\theta_L,\varnothing},\varnothing) - W(x_{\theta_H,\varnothing},\varnothing))) + \frac{1}{2}(e_{\theta_H,\varnothing} + e_{\theta_L,\varnothing} - W(x_{\theta_L,\varnothing},\varnothing)\\ -W(x_{\theta_H,\varnothing},\varnothing)) - \Pi_{\varnothing M} + \mu\frac{1}{2}p'(\mu_1)((e_{\theta_L,\theta_L} + e_{\theta_H,\theta_H} - W(x_{\theta_L,\theta_L},\theta_L)\\ -W(x_{\theta_H,\theta_H},\theta_H) - S(x_{\theta_H,\theta_H},\theta_H) - (e_{\theta_H,\varnothing} + e_{\theta_L,\varnothing} - W(x_{\theta_L,\varnothing},\varnothing) - W(x_{\theta_H,\varnothing},\varnothing))) + \varphi_2 = 0,\end{aligned}$$

where φ_2 is the Lagrange multiplier of (PCM). Assume that only the (CC1) binds while the Participation Constraint of the monitor is slack. Then $\varphi_2 = 0$. Substituting by the values obtained in Proposition 4.1, we can rewrite the first order condition as,

$$\frac{1}{2}p(\mu)\frac{\Delta\theta^2}{2(1-p(\mu))^2} - \frac{1}{4}\frac{\Delta\theta^2}{2(1-p(\mu))^2}p(\mu)^2 + \mu_1\frac{1}{2}p'(\mu)\frac{\Delta\theta^2}{2(1-p(\mu))^2} = 0. \tag{36}$$

Which can be further simplified to

$$\frac{\Delta\theta^2}{4(1-p(\mu))^2}p(\mu)\left(1 - p_0e^{-\lambda\mu} - \lambda\mu\right) = 0$$

Notice that if $\mu_1 = 0$ that equation is positive, hence $\mu_1 = 0$ is never a solution to the firm design. In particular, the solution is implicitly given by

$$\mu = \frac{1 - p_0 e^{-\lambda \mu_1}}{\lambda}$$

Rearraging and defining $u = 1 - \lambda \mu_1$,

$$-ue^{-u} = -p_0 e^{-1},$$

which is the standard form of the Lambert W, which means that $u = W(-p_0 e^{-1})$. Substituting again u, we get

$$\mu = \frac{1 + W_0\left(-p_0 e^{-1}\right)}{\lambda}$$

Going back to the Participation Constraint of the Monitor, we can substitute the values of the contract in (PCM) and define $\bar{V}_2$ as,

$$\bar{V}_2 := -W_0\left(-p_0 e^{-1}\right) \cdot \frac{1 + W_0\left(-p_0 e^{-1}\right)}{\lambda} \cdot \frac{\Delta\theta}{2} \cdot \left(1 - \Delta\theta \frac{3 - p(\mu)}{2(1 - p(\mu))}\right)$$

□

Proof of Corollary 4.1

Proof. First, it is important to know that the main branch of the Lambert W function is an increasing function with

$$W_0\left(-\frac{1}{e}\right) = -1, \; W_0(0) = 0$$

This guarantees that $1 + W_0\left(-p_0 e^{-1}\right)$ is always positive. As $W_0\left(-p_0 e^{-1}\right) < 0$, we have that

$$\frac{1 + W_0\left(-\frac{p_0}{e}\right)}{\lambda} < \frac{1}{\lambda}$$

which guarantees that the optimal span of control under collusion is always smaller than in the non collusion case. At the same time, if $\mu_1 \in (0, 1)$, applying the implicit function theorem and the chain rule, we can see that

$$\frac{\partial \mu_1}{\partial \lambda} = -\frac{1 + W_0\left(-p_0 e^{-1}\right)}{\lambda^2} < 0$$
$$\frac{\partial \mu_1}{\partial p_0} = -\frac{W_0'\left(-p_0 e^{-1}\right)}{\lambda e} < 0$$

where the second inequality comes from the fact that W_0 is an increasing function. Finally, the interior optimal condition is not satisfied if

$$\frac{1 + W_0\left(-p_0 e^{-1}\right)}{\lambda} \geq 1$$

or equivalently,

$$\lambda \leq 1 + W_0\left(-p_0 e^{-1}\right)$$

In this case, the optimal $\mu_1 = 1$. □

Proof of Proposition 4.3 If only the participation constraint is binding, the problem of the firm can be solved as in the non-collusion case, and therefore, the expected cost of a monitor is $\bar{V}$. Else, if the participation constraint is not binding, the problem under one monitor and under two monitors are unrelated, as they do not share any restriction in common. Hence, both problems can be solved separately. Notice that the solution to the problem of one monitor is given by Proposition 4.1. I will rewrite the problem taking into account only the problem of two monitors and assuming that the participation constraint is not binding.

$$
\begin{aligned}
\max_{\{S,W,e\}} \; & \left(1-(1-p(\mu_1))^2\right)\left(e_{1,2}+e_{4,2}-W(x_{\theta_L,\theta_L},\theta_L)-W(x_{\theta_H,\theta_H},\theta_H)\right) \\
& + (1-p(\mu_1))^2\left(e_{2,2}+e_{3,2}-W(x_{\theta_L,\varnothing},\varnothing)-W(x_{\theta_H,\varnothing},\varnothing)\right) \\
& - 2p(\mu_1)(S(x_{\theta_L,\theta_L},\theta_L)+S(x_{\theta_H,\theta_H},\theta_H)) \\
& - 2(1-p(\mu_1))(S(x_{\theta_L,\varnothing},\varnothing)+S(x_{\theta_H,\varnothing},\varnothing)) - p(\mu_1)(1-p(\mu_1))(\delta-\beta)
\end{aligned}
$$

s.t.

$$S(x_i,r_i)\geq 0,\;\; i=1,2,4 \tag{LLS}$$

$$S^2(x_i,r_i)-\beta\geq 0,\;\; i=3 \tag{LLS'}$$

$$\beta\geq 0, \tag{PenC}$$

$$W^2(x_i,r_i)-\frac{e_i^2}{2}\geq 0,\;\; i=1,2,3,4 \tag{IC1'}$$

$$W(x_{\theta_H,\varnothing},\varnothing)-\frac{e_{3,2}^2}{2}\geq W(x_{\theta_L,\varnothing},\varnothing)-\frac{(e_{2,2}-\Delta\theta)^2}{2} \tag{IC3'}$$

$$
\begin{aligned}
&2S(x_{\theta_H,\theta_H},\theta_H)+W(x_{\theta_H,\theta_H},\theta_H)-c(e_{4,1})\geq \\
&2S(x_{\theta_H,\varnothing},\varnothing)+W(x_{\theta_H,\varnothing},\varnothing)-c(e_{3,1})
\end{aligned} \tag{CC3}
$$

$$
\begin{aligned}
&S(x_{\theta_H,\theta_H},\theta_H)+\delta-\beta+W(x_{\theta_H,\theta_H},\theta_H)-c(e_{4,1})\geq \\
&S(x_{\theta_H,\varnothing},\varnothing)+W(x_{\theta_H,\varnothing},\varnothing)-c(e_{3,1})
\end{aligned} \tag{CC4}
$$

$$S(x_{\theta_H,\theta_H},\theta_H)\geq S(x_{\theta_H,\varnothing},\varnothing)+\delta-\beta \tag{CC5}$$

The Lagrangian is

$$
\begin{aligned}
\mathcal{L}() = & \left(1-(1-p(\mu_1))^2\right)\left(e_{1,2}+e_{4,2}-W(x_{\theta_L,\theta_L},\theta_L)-W(x_{\theta_H,\theta_H},\theta_H)\right) \\
& + (1-p(\mu_1))^2\left(e_{2,2}+e_{3,2}-W(x_{\theta_L,\varnothing},\varnothing)-W(x_{\theta_H,\varnothing},\varnothing)\right) \\
& - 2p(\mu_1)(S(x_{\theta_L,\theta_L},\theta_L)+S(x_{\theta_H,\theta_H},\theta_H)) \\
& - 2(1-p(\mu_1))(S(x_{\theta_L,\varnothing},\varnothing)+S(x_{\theta_H,\varnothing},\varnothing)) - p(\mu_1)(1-p(\mu_1))(\delta-\beta) \\
& + \sum_{i=1,2,4}\lambda_i\left(S(x_i,r_i)\right)+\lambda_3(S_3-\beta)+\gamma\beta+\sum_{i=1,2,4}\varphi_i\left(W_i-\frac{e_i^2}{2}\right) \\
& + \varphi_3\left(W_3-\frac{e_3^2}{2}-W_2+\frac{(e_2-\Delta\theta)^2}{2}\right)+\Pi_3\left(2S_4+W_4-\frac{e_4^2}{2}-2S_3-W_3+\frac{e_3^2}{2}\right) \\
& \Pi_4\left(S_4+\delta-\beta+W_4-\frac{e_4^2}{2}-S_3-W_3+\frac{e_3^2}{2}\right)+\Pi_5\left(S_4-S_3-\delta+\beta\right)
\end{aligned}
$$

And the first order conditions of the problem are:

$$\frac{\partial \mathcal{L}}{\partial e_1} = 0 \Rightarrow (1-(1-p(\mu_1))^2) - \varphi_1 e_1 = 0 \tag{37}$$
$$\frac{\partial \mathcal{L}}{\partial e_2} = 0 \Rightarrow (1-p(\mu_1))^2 - \varphi_2 e_2 + \varphi_3(e_2 - \Delta\theta) = 0 \tag{38}$$
$$\frac{\partial \mathcal{L}}{\partial e_3} = 0 \Rightarrow (1-p(\mu_1))^2 - \varphi_3 e_3 + \Pi_3 e_3 + \Pi_4 e_3 = 0 \tag{39}$$
$$\frac{\partial \mathcal{L}}{\partial e_4} = 0 \Rightarrow (1-(1-p(\mu_1))^2) - \varphi_4 e_4 - \Pi_3 e_4 - \Pi_4 e_4 = 0 \tag{40}$$
$$\frac{\partial \mathcal{L}}{\partial W_1} = 0 \Rightarrow (1-(1-p(\mu_1))^2) - \varphi_1 = 0 \tag{41}$$
$$\frac{\partial \mathcal{L}}{\partial W_2} = 0 \Rightarrow (1-p(\mu_1))^2 - \varphi_2 + \varphi_3 = 0 \tag{42}$$
$$\frac{\partial \mathcal{L}}{\partial W_3} = 0 \Rightarrow (1-p(\mu_1))^2 - \varphi_3 + \Pi_3 + \Pi_4 = 0 \tag{43}$$
$$\frac{\partial \mathcal{L}}{\partial W_4} = 0 \Rightarrow (1-(1-p(\mu_1))^2) - \varphi_4 - \Pi_3 - \Pi_4 = 0 \tag{44}$$
$$\frac{\partial \mathcal{L}}{\partial S_1} = 0 \Rightarrow -2p(\mu_1) + \lambda_1 = 0 \tag{45}$$
$$\frac{\partial \mathcal{L}}{\partial S_2} = 0 \Rightarrow -2(1-p(\mu_1)) + \lambda_2 = 0 \tag{46}$$
$$\frac{\partial \mathcal{L}}{\partial S_3} = 0 \Rightarrow -2(1-p(\mu_1)) + \lambda_3 - 2\Pi_3 - \Pi_4 - \Pi_5 = 0 \tag{47}$$
$$\frac{\partial \mathcal{L}}{\partial S_4} = 0 \Rightarrow -2p(\mu_1) + \lambda_4 + 2\Pi_3 + \Pi_4 + \Pi_5 = 0 \tag{48}$$
$$\frac{\partial \mathcal{L}}{\partial \delta} = 0 \Rightarrow -p(\mu_1)(1-p(\mu_1)) + \Pi_4 - \Pi_5 = 0 \tag{49}$$
$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \Rightarrow p(\mu_1)(1-p(\mu_1)) + \gamma - \lambda_3 - \Pi_4 + \Pi_5 = 0 \tag{50}$$

As in previous proofs, it is immediate to see that $e_{1,2} = e_{3,2} = e_{4,2} = 1$. From the first order condition of $W(x_{\theta_L,\theta_L}, \theta_L)$, we can see that $\varphi_1 > 0$, which means that $W_1 = \frac{1}{2}$. At the same time, form (42), $\varphi_2 > 0$, which also means that $W_2 = \frac{e_2^2}{2}$. So the low type agents do not receive any rents under the optimal contract. From equation (43), $\varphi_3 > 0$. This means that the (IC3) is binding and,

$$W_3 = \frac{1}{2} + \frac{e_2^2}{2} - \frac{(e_2 - \Delta\theta)^2}{2} = \frac{1-\Delta\theta^2}{2} + e_2\Delta\theta$$

From equations (45) and (46) it is immediate that $\lambda_1, \lambda_2 > 0$, which implies that $S_1 = S_2 = 0$. From equation (47),

$$\lambda_3 = 2(1-p(\mu_1)) + 2\Pi_3 + \Pi_4 + \Pi_5 > 0,$$

which implies that $S_3 = 0$. This implies that, when the *Participation Constraint of the Monitor* is not binding, she only receives wages from reporting high types.

From the first order condition of δ, we have that $\Pi_4 > 0$, which means that (CC4) is binding. Adding (49) and (50), we get $\gamma = \lambda_3$, which implies that $\gamma > 0$ and therefore $\beta = 0$.

If $S_4 = \delta$, equation (CC3) becomes (CC4), which we already know is binding. If, instead $S_4 > \delta$, $\Pi_5 = 0$ and (CC3) is not binding, as

$$2S_4 + W_4 > S_4 + \delta + W_4 = W_3.$$

This implies that $\Pi_3 = 0$. From (48) we have that $\Pi_4 + \lambda_4 = 2p(\mu_1)$, and from (50) $\Pi_4 = p(\mu_1)(1 - p(\mu_1))$. If $S_4 > 0$, this is a contradiction. Now, if $S_4 = 0$, then $S_4 = \delta = 0$, which makes (CC5) binding and thus, we arrive to a contradiction.

To finish the proof, we just need to compute e_2 and W_4. As (CC4) is binding, we know that

$$W_4 + 2S_4 = W_3.$$

Notice that any combination of S_4 and W_4 that satisfies the restrictions and the previous equation would give the same profits for the monitor. This implies that $\varphi_4 = 0$. As in previous proofs, I will assume that $W_4 = \frac{1}{2}$, satisfying the incentive compatibility constraint, and $2S_4 = W_3 - \frac{1}{2}$.

Adding equations (43) and (44) we get that $\varphi_3 = 1$. And from equation (42) $\varphi_3 - \varphi_2 = (1 - p(\mu_1))^2$. Substituting into the first order condition of e_2,

$$e_2 = 1 - \frac{\Delta\theta}{(1 - p(\mu_1))^2},$$

finishing the proof.

Proof of Proposition 4.5.

Proof. As (PCM) is never binding, Proposition 4.4 guarantees that $\mu_1 = \mu^2$. Hence, the profit function can be written as,

$$\mathbb{E}\left[\Pi(\Omega, \mathbf{W}, \mathbf{S})\right] = \Pi_{\varnothing M} + \frac{\Delta\theta^2(1-(1-p(\mu_1)^2))}{4(1-p(\mu_1))^2}\mu_1$$

In this case,

$$\begin{aligned} \mu_1 = & \frac{(1-p_0e^{-\lambda\mu_1}) + (1-p_0e^{-\lambda\mu_1})^2}{2\lambda} \\ = & \frac{(1-p_0e^{-\lambda\mu_1})(2-p_0e^{-\lambda\mu_1})}{2\lambda} \end{aligned}$$

To show the uniqueness, assume that μ_1 is interior (as otherwise is trivial) and define $\phi(\mu) := \frac{(1-p(\mu))+(1-p(\mu))^2}{2\lambda}$. One checks

$$\phi'(\mu) = \frac{p(\mu)\left[3-2p(\mu)\right]}{2} \in (0, \tfrac{9}{16}) \qquad \text{for } p(\mu) \in (0,1),$$

so ϕ is a contraction. Then, Banach theorem ensures that there is a unique interior fixed point $\mu_1 = \phi(\mu_1)$. If that fixed point exceeds 1 (small λ or small p_0), the optimum hits the boundary $\mu_1 = 1$.

□

Proof of Proposition 4.6

Proof. Define

$$F(\mu_1, p_0, \lambda) \; := \; \mu_1 - \frac{g(p(\mu_1))}{2\lambda}, \qquad g(p) := (1-p)(2-p) = 2-3p+p^2, \qquad p(\mu_1) = p_0e^{-\lambda\mu_1}.$$

An interior optimum μ_1^* satisfies $F(\mu_1^*, p_0, \lambda) = 0$. Compute the needed partial derivatives. First,

$$\frac{\partial F}{\partial \mu_1} = 1 - \frac{g'(p)}{2\lambda}\frac{\partial p}{\partial \mu_1} = 1 - \frac{(-3+2p)(-\lambda p)}{2\lambda} = 1 - \frac{p(3-2p)}{2} = \frac{2-3p+2p^2}{2}.$$

For $p \in (0,1)$ the quadratic $2-3p+2p^2 > 0$, hence $\partial F/\partial\mu_1 > 0$. By the Implicit Function Theorem, there is a C^1 function $\mu_1^*(p_0, \lambda)$ solving $F = 0$, and

$$\frac{\partial \mu_1^*}{\partial x} = - \left.\frac{\partial F/\partial x}{\partial F/\partial \mu_1}\right|_{F=0}, \qquad x \in \{p_0, \lambda\}.$$

Derivative with respect to p_0. Holding μ_1 fixed,

$$\frac{\partial F}{\partial p_0} = -\frac{g'(p)}{2\lambda}\frac{\partial p}{\partial p_0} = -\frac{(-3+2p)}{2\lambda}e^{-\lambda\mu_1} = \frac{(3-2p)p}{2\lambda p_0}.$$

Therefore

$$\frac{\partial \mu_1^*}{\partial p_0} = -\frac{\frac{(3-2p)p}{2\lambda p_0}}{\frac{2-3p+2p^2}{2}} = -\frac{p(3-2p)}{\lambda p_0(2-3p+2p^2)} < 0,$$

since $p > 0$, $3-2p > 0$ and the denominator is positive.

Derivative with respect to λ. Holding μ_1 fixed,

$$\frac{\partial F}{\partial \lambda} = -\frac{\partial}{\partial \lambda}\left(\frac{g(p)}{2\lambda}\right) = \frac{g(p)}{2\lambda^2} - \frac{g'(p)}{2\lambda}\frac{\partial p}{\partial \lambda} = \frac{g(p)}{2\lambda^2} - \frac{(-3+2p)}{2\lambda}(-\mu_1 p).$$

Using the first order condition, $\mu_1 = \frac{g(p)}{2\lambda}$, gives

$$\frac{\partial F}{\partial \lambda} = \frac{g(p)}{2\lambda^2} - \frac{g(p)p(3-2p)}{4\lambda^2} = \frac{g(p)}{4\lambda^2}\big(2-p(3-2p)\big) = \frac{g(p)}{2\lambda^2}\frac{\partial F}{\partial \mu_1}.$$

Hence

$$\frac{\partial \mu_1^*}{\partial \lambda} = -\frac{\frac{g(p)}{2\lambda^2}\frac{\partial F}{\partial \mu_1}}{\frac{\partial F}{\partial \mu_1}} = -\frac{g(p)}{2\lambda^2} = -\frac{(1-p)(2-p)}{2\lambda^2} < 0,$$

because $(1-p)(2-p) > 0$ for $p \in (0,1)$. □

Proof of proposition A.1

Proof. Throughout write $K := (1-\Delta\theta/2)^2$ and $p := p(\mu)$.

Step 1 (FOC and derivatives). For an interior maximizer,

$$F(\mu;\lambda) := \frac{\partial \Delta\theta\Pi(\mu)}{\partial \mu} = 0.$$

A direct differentiation yields

$$F(\mu;\lambda) = \underbrace{\tfrac{1}{2}\Big[\Delta\theta(1-\Delta\theta) - \tfrac{1}{2}(1-p)K\Big] - \tfrac{1}{4}\lambda\mu K p}_{=:F_A} + \underbrace{\lambda\frac{(1/\Delta\theta - 3/2)}{p}\bar{V}}_{=:F_B} + \underbrace{\left(\frac{1-p}{\Delta\theta^2\mu^2p^2} - \frac{\lambda}{\Delta\theta^2\mu}\frac{2-p}{p^2}\right)\bar{V}^2}_{=:F_C}. \tag{51}$$

The second derivative $F_\mu = \partial^2\Delta\theta\Pi/\partial\mu^2$ is negative at the interior optimum (SOC), so by the Implicit Function Theorem

$$\frac{d\mu^*}{d\lambda} = -\frac{F_\lambda}{F_\mu}\bigg|_{\mu=\mu^*}, \qquad \text{hence} \quad \text{sign}\left(\frac{d\mu^*}{d\lambda}\right) = \text{sign}(F_\lambda) \quad \text{at } \mu^*.$$

Step 2 (Asymptotics $\lambda \to \infty$). Let $x := \lambda\mu \in (0, \lambda]$, so $p(\mu) = p_0 e^{-x}$. Rewrite $\Delta\theta\Pi$ grouping orders in λ:

$$\Delta\theta\Pi(\mu) = \underbrace{\frac{x}{2\lambda}\Big[\Delta\theta(1-\Delta\theta) - \tfrac{1}{2}\big(1 - p_0 e^{-x}\big)K\Big]}_{O(1/\lambda)} + \underbrace{\Big(\tfrac{1}{\Delta\theta p_0 e^{-x}} - \tfrac{3}{2p_0 e^{-x}} + \tfrac{1}{2} - \tfrac{1}{\Delta}\theta\Big)\bar{V}}_{O(1)} - \underbrace{\lambda\frac{e^{2x} - p_0 e^{x}}{\Delta\theta^2 x p_0^2}}_{=:\ \lambda g(x) > 0}\bar{V}^2.$$

As $\lambda \to \infty$, the dominant term is $-\lambda g(x)\bar{V}^2$, with g continuous on $(0, \infty)$ and $g(x) \to \infty$ as $x \downarrow 0$ or $x \to \infty$. Hence g attains a finite minimum at some $x^* \in (0, \infty)$, and the maximizer of $\Delta\theta\Pi$ must choose x near x^*, i.e.

$$\mu^*(\lambda) = \frac{x^*}{\lambda} + o\Big(\frac{1}{\lambda}\Big) \quad \Longrightarrow \quad \mu^*(\lambda) \to 0,$$

which proves part (i).

Step 3 (Sign of $d\mu^*/d\lambda$ for moderate λ). Differentiate F in (51) w.r.t. λ at an interior optimum. Using $p_\lambda = -\mu p$ and collecting terms,

$$F_\lambda = \underbrace{-\tfrac{1}{4}\mu^* K p + \tfrac{1}{4}\lambda\mu^{*2} K p}_{\text{from } F_A} + \underbrace{\lambda\frac{(1/\Delta\theta - 3/2)}{p}\bar{V}}_{\text{from } F_B} + \underbrace{\frac{\bar{V}^2}{\Delta\theta^2}\Big(\frac{1}{\mu^{*2}}\frac{1-p}{p^2} - \frac{\lambda}{\mu^*}\frac{2-p}{p^2}\Big)}_{\text{from } F_C}.$$

For $\Delta\theta < \frac{2}{3}$ one has $(1/\Delta\theta - 3/2) > 0$. On any compact set of (λ, μ) where the CC + PCM solution is interior and p stays away from 0 and 1, the (positive) middle term $\lambda\frac{(1/\Delta\theta - 3/2)}{p}\bar{V}$ dominates the negative contribution from the $\bar{V}^2$–term for all *moderate* λ and $\bar{V}$ small enough. Hence there exists $\lambda_0 > 0$ (within the regime) such that $F_\lambda(\mu^*(\lambda_0); \lambda_0) > 0$, and therefore $d\mu^*/d\lambda > 0$ at λ_0.

Step 4 (Sign of $d\mu^*/d\lambda$ for large λ). From Step 2, $\mu^*(\lambda) = x^*/\lambda + o(1/\lambda)$ and $p(\mu^*) = p_0 e^{-x^*} \in (0, p_0]$ as $\lambda \to \infty$. Substituting this scaling into F_λ and keeping the leading order gives

$$F_\lambda(\mu^*; \lambda) \;=\; -\frac{\lambda}{\Delta\theta^2}\frac{2 - p_0 e^{-x^*}}{p_0^2 e^{-2x^*}} \cdot \frac{\bar{V}^2}{x^*} \;+\; O(1),$$

which is strictly negative for all sufficiently large λ. Since $F_\mu < 0$ at the optimum, we conclude $d\mu^*/d\lambda < 0$ eventually.

Step 5 (Non-monotonicity). By Steps 3–4 there exist λ_0 with $d\mu^*/d\lambda > 0$ and $\Lambda > \lambda_0$ with $d\mu^*/d\lambda < 0$ (both within the CC + PCM regime). Continuity of $\mu^*(\cdot)$ and of F_λ implies at least one turning point $\lambda^* \in (\lambda_0, \Lambda)$ where $d\mu^*/d\lambda = 0$, hence $\mu^*(\lambda)$ is non-monotonic. This proves part (ii). □

Proof of Proposition B.1

Proof. For this proof assume that $\mu_1 > \mu_2$. In the opposite case, the optimal contract can be solved as in Proposition 4.1 with a different probability of detection. For this part, I will assume that both the (PCM) and at least a collusion constraint are binding, as if all the Collusion Constraints are non-binding, we are exactly in the non-collusion scenario discussed in the previous section. So, at least, one collusion constraint must be binding. Take again the maximization problem of the principal. Now, as (PCM) is binding, we can not split the two problems of the monitor. For completeness I will write the maximization problem of the monitor, the Lagrangian and its first order conditions.

$$
\begin{aligned}
\max_{\{S,W,e\}} \ & (1-2\mu_1+\mu_2)\Pi_{\varnothing M} + (\mu_1-\mu_2)p(\mu_1)(e_{1,1}+e_{4,1}-W^1(x_{\theta_L,\theta_L},\theta_L)-W^1(x_{\theta_H,\theta_H},\theta_H)) \\
& + (\mu_1-\mu_2)(1-p(\mu_1))(e_{2,1}+e_{3,1}-W^1(x_{\theta_L,\varnothing},\varnothing)-W^1(x_{\theta_H,\varnothing},\varnothing)) \\
& + \frac{1}{2}\mu_2(1-(1-p(\mu_1))^2)\left(e_{1,2}+e_{4,2}-W^2(x_{\theta_L,\theta_L},\theta_L)-W^2(x_{\theta_H,\theta_H},\theta_H)\right) \\
& + \frac{1}{2}\mu_2(1-p(\mu_1))^2\left(e_{2,2}+e_{3,2}-W^2(x_{\theta_L,\varnothing},\varnothing)-W^2(x_{\theta_H,\varnothing},\varnothing)\right) \\
& - (\mu_1-\mu_2)\left(p(\mu_1)(S^1(x_{\theta_L,\theta_L},\theta_L)+S^1(x_{\theta_H,\theta_H},\theta_H))+(1-p(\mu_1))(S^1(x_{\theta_L,\varnothing},\varnothing)+S^1(x_{\theta_H,\varnothing},\varnothing))\right) \\
& - \mu_2\left[p(\mu_1)(S^2(x_{\theta_L,\theta_L},\theta_L)+S^2(x_{\theta_H,\theta_H},\theta_H))+(1-p(\mu_1))(S^2(x_{\theta_L,\varnothing},\varnothing)+S^2(x_{\theta_H,\varnothing},\varnothing))\right. \\
& \left. + p(\mu_1)(1-p(\mu_1))(\delta-\beta)\right]
\end{aligned}
$$

s.t.

$$S^1(x_i,r_i) \geq 0, \ i=1,2,3,4 \tag{LLS}$$

$$S^2(x_i,r_i) \geq 0, \ i=1,2,4 \tag{LLS}$$

$$S^2(x_3,r_3)-\beta \geq 0, \tag{LLS'}$$

$$\beta \geq 0, \tag{PenC}$$

$$W^1(x_i,r_i)-\frac{e_{i,1}^2}{2} \geq 0, \ i=1,2,3,4 \tag{IC1}$$

$$W^2(x_i,r_i)-\frac{e_{i,2}^2}{2} \geq 0, \ i=1,2,3,4 \tag{IC1'}$$

$$W^1(x_{\theta_H,\varnothing},\varnothing)-\frac{e_{3,1}^2}{2} \geq W^1(x_{\theta_L,\varnothing},\varnothing)-\frac{(e_{2,1}-\Delta\theta)^2}{2} \tag{IC3}$$

$$W^2(x_{\theta_H,\varnothing},\varnothing)-\frac{e_{3,2}^2}{2} \geq W^2(x_{\theta_L,\varnothing},\varnothing)-\frac{(e_{2,2}-\Delta\theta)^2}{2} \tag{IC3'}$$

$$
\begin{aligned}
& (\mu_1-\mu_2)\left(p(\mu_1)(S^1(x_{\theta_L,\theta_L},\theta_L)+S^1(x_{\theta_H,\theta_H},\theta_H))+(1-p(\mu_1))(S^1(x_{\theta_L,\varnothing},\varnothing)+S^1(x_{\theta_H,\varnothing},\varnothing))\right) \\
& + \mu_2\left(p(\mu_1)(S^2(x_{\theta_L,\theta_L},\theta_L)+S^2(x_{\theta_H,\theta_H},\theta_H))+(1-p(\mu_1))(S^2(x_{\theta_L,\varnothing},\varnothing)+S^2(x_{\theta_H,\varnothing},\varnothing))\right. \\
& \left. + p(\mu_1)(1-p(\mu_1))(\delta-\beta)\right) \geq 2\bar{V}
\end{aligned} \tag{PCM}
$$

$$
\begin{aligned}
& S^1(x_{\theta_H,\theta_H},\theta_H)+W^1(x_{\theta_H,\theta_H},\theta_H)-c(e_{4,1}) \geq \\
& S^1(x_{\theta_H,\varnothing},\varnothing)+W^1(x_{\theta_H,\varnothing},\varnothing)-c(e_{3,1})
\end{aligned} \tag{CC1}
$$

$$
\begin{aligned}
& 2S^2(x_{\theta_H,\theta_H},\theta_H)+W^2(x_{\theta_H,\theta_H},\theta_H)-c(e_{4,2}) \geq \\
& 2S^2(x_{\theta_H,\varnothing},\varnothing)+W^2(x_{\theta_H,\varnothing},\varnothing)-c(e_{3,2})
\end{aligned} \tag{CC3}
$$

$$
\begin{aligned}
& S^2(x_{\theta_H,\theta_H},\theta_H)+\delta-\beta+W^2(x_{\theta_H,\theta_H},\theta_H)-c(e_{4,2}) \geq \\
& S^2(x_{\theta_H,\varnothing},\varnothing)+W^2(x_{\theta_H,\varnothing},\varnothing)-c(e_{3,2})
\end{aligned} \tag{CC4}
$$

$$S^2(x_{\theta_H,\theta_H},\theta_H) \geq S^2(x_{\theta_H,\varnothing},\varnothing)+\delta-\beta \tag{CC5}$$

The Lagrangian.

$$
\begin{aligned}
\mathcal{L}() = {} & (1-2\mu_1+\mu_2)\Pi_{\varnothing M} + (\mu_1-\mu_2)p(\mu_1)(e_{1,1}+e_{4,1}-W^1(x_{\theta_L,\theta_L},\theta_L)-W^1(x_{\theta_H,\theta_H},\theta_H)) \\
& + (\mu_1-\mu_2)(1-p(\mu_1))(e_{2,1}+e_{3,1}-W^1(x_{\theta_L,\varnothing},\varnothing)-W^1(x_{\theta_H,\varnothing},\varnothing)) \\
& + \frac{1}{2}\mu_2(1-(1-p(\mu_1))^2)\left(e_{1,2}+e_{4,2}-W^2(x_{\theta_L,\theta_L},\theta_L)-W^2(x_{\theta_H,\theta_H},\theta_H)\right) \\
& + \frac{1}{2}\mu_2(1-p(\mu_1))^2\left(e_{2,2}+e_{3,2}-W^2(x_{\theta_L,\varnothing},\varnothing)-W^2(x_{\theta_H,\varnothing},\varnothing)\right) \\
& - (\mu_1-\mu_2)\left(p(\mu_1)(S^1(x_{\theta_L,\theta_L},\theta_L)+S^1(x_{\theta_H,\theta_H},\theta_H))+(1-p(\mu_1))(S^1(x_{\theta_L,\varnothing},\varnothing)+S^1(x_{\theta_H,\varnothing},\varnothing))\right) \\
& - \mu_2\left[p(\mu_1)(S^2(x_{\theta_L,\theta_L},\theta_L)+S^2(x_{\theta_H,\theta_H},\theta_H))+(1-p(\mu_1))(S^2(x_{\theta_L,\varnothing},\varnothing)+S^2(x_{\theta_H,\varnothing},\varnothing))\right. \\
& \left. +p(\mu_1)(1-p(\mu_1))(\delta-\beta)\right] \\
& + \sum_{i=1,2,3,4}\lambda_i^1\left(S^1(x_i,r_i)\right) + \sum_{i=1,2,4}\lambda_i^2\left(S^2(x_i,r_i)\right) + \lambda_3^2(S^2(x_{\theta_H,\varnothing},\varnothing)-\beta)+\gamma\beta \\
& + \sum_{i=1,2,4}\varphi_i^1\left(W_i^1-\frac{e_{i,1}^2}{2}\right) + \sum_{i=1,2,4}\varphi_i^2\left(W_i^2-\frac{e_{i,2}^2}{2}\right) \\
& + \varphi_3^1\left(W^1(x_{\theta_H,\varnothing},\varnothing)-\frac{e_{3,1}^2}{2}-W^1(x_{\theta_L,\varnothing},\varnothing)+\frac{(e_{2,1}-\Delta\theta)^2}{2}\right) \\
& + \varphi_3^2\left(W^2(x_{\theta_H,\varnothing},\varnothing)-\frac{e_{3,2}^2}{2}W^2(x_{\theta_L,\varnothing},\varnothing)+\frac{(e_{2,2}-\Delta\theta)^2}{2}\right) \\
& + \Pi_1\left(S^1(x_{\theta_H,\theta_H},\theta_H)+W^1(x_{\theta_H,\theta_H},\theta_H)-\frac{e_{4,1}^2}{2}-S^1(x_{\theta_H,\varnothing},\varnothing)-W^1(x_{\theta_H,\varnothing},\varnothing)+\frac{e_{3,1}^2}{2}\right) \\
& + \Pi_3\left(2S^2(x_{\theta_H,\theta_H},\theta_H)+W_4^2-\frac{e_{4,2}^2}{2}-2S^2(x_{\theta_H,\varnothing},\varnothing)-W^2(x_{\theta_H,\varnothing},\varnothing)+\frac{e_{3,2}^2}{2}\right) \\
& + \Pi_4\left(S^2(x_{\theta_H,\theta_H},\theta_H)+\delta-\beta+W_4^2-\frac{e_{4,2}^2}{2}-S^2(x_{\theta_H,\varnothing},\varnothing)-W^2(x_{\theta_H,\varnothing},\varnothing)+\frac{e_{3,2}^2}{2}\right)+\Pi_5\big(S^2(x_{\theta_H,\theta_H},\theta \\
& + \alpha\left((\mu_1-\mu_2)\left(p(\mu_1)(S^1(x_{\theta_L,\theta_L},\theta_L)+S^1(x_{\theta_H,\theta_H},\theta_H))+(1-p(\mu_1))(S(x_{\theta_L,\varnothing},\varnothing)^1+S^1(x_{\theta_H,\varnothing},\varnothing))\right)\right. \\
& +\mu_2\left(p(\mu_1)(S^2(x_{\theta_L,\theta_L},\theta_L)+S^2(x_{\theta_H,\theta_H},\theta_H))+(1-p(\mu_1))(S(x_{\theta_L,\varnothing},\varnothing)^2+S^2(x_{\theta_H,\varnothing},\varnothing))\right. \\
& \left.\left. +p(\mu_1)(1-p(\mu_1))(\delta-\beta)\right)-2\bar{V}\right)
\end{aligned}
$$

First order conditions.

$$\frac{\partial \mathcal{L}}{\partial e_{1,1}} = 0 \Rightarrow (\mu_1 - \mu_2)p(\mu_1) - \varphi_1^1 e_{1,1} = 0 \tag{52}$$

$$\frac{\partial \mathcal{L}}{\partial e_{2,1}} = 0 \Rightarrow (\mu_1 - \mu_2)(1 - p(\mu_1)) - \varphi_2^1 e_{2,1} + \varphi_3^1(e_{2,1} - \Delta\theta) = 0 \tag{53}$$

$$\frac{\partial \mathcal{L}}{\partial e_{3,1}} = 0 \Rightarrow (\mu_1 - \mu_2)(1 - p(\mu_1)) - \varphi_3^1 e_{3,1} + \Pi_1 e_{3,1} = 0 \tag{54}$$

$$\frac{\partial \mathcal{L}}{\partial e_{4,1}} = 0 \Rightarrow (\mu_1 - \mu_2)p(\mu_1) - \varphi_4^1 e_{4,1} - \Pi_1 e_{4,1} = 0 \tag{55}$$

$$\frac{\partial \mathcal{L}}{\partial e_{1,2}} = 0 \Rightarrow \frac{1}{2}\mu_2(1 - (1 - p(\mu_1))^2) - \varphi_1^2 e_{1,2} = 0 \tag{56}$$

$$\frac{\partial \mathcal{L}}{\partial e_{2,2}} = 0 \Rightarrow \frac{1}{2}\mu_2(1 - p(\mu_1))^2 - \varphi_2^2 e_{2,2} + \varphi_3^2(e_{2,2} - \Delta\theta) = 0 \tag{57}$$

$$\frac{\partial \mathcal{L}}{\partial e_{3,2}} = 0 \Rightarrow \frac{1}{2}\mu_2(1 - p(\mu_1))^2 - \varphi_3^2 e_{3,2} + \Pi_3 e_{3,2} + \Pi_4 e_{3,2} = 0 \tag{58}$$

$$\frac{\partial \mathcal{L}}{\partial e_{4,2}} = 0 \Rightarrow \frac{1}{2}\mu_2(1 - (1 - p(\mu_1))^2) - \varphi_4^2 e_{4,2} - \Pi_3 e_{4,2} - \Pi_4 e_{4,2} = 0 \tag{59}$$

$$\frac{\partial \mathcal{L}}{\partial W^1(x_{\theta_L,\theta_L}, \theta_L)} = 0 \Rightarrow (\mu_1 - \mu_2)p(\mu_1) - \varphi_1^1 = 0 \tag{60}$$

$$\frac{\partial \mathcal{L}}{\partial W^1(x_{\theta_L,\varnothing}, \varnothing)} = 0 \Rightarrow (\mu_1 - \mu_2)(1 - p(\mu_1)) - \varphi_2^1 + \varphi_3^1 = 0 \tag{61}$$

$$\frac{\partial \mathcal{L}}{\partial W^1(x_{\theta_H,\varnothing}, \varnothing)} = 0 \Rightarrow (\mu_1 - \mu_2)(1 - p(\mu_1)) - \varphi_3^1 + \Pi_1 = 0 \tag{62}$$

$$\frac{\partial \mathcal{L}}{\partial W^1(x_{\theta_H,\theta_H}, \theta_H)} = 0 \Rightarrow (\mu_1 - \mu_2)p(\mu_1) - \varphi_4^1 - \Pi_1 = 0 \tag{63}$$

$$\frac{\partial \mathcal{L}}{\partial W^2(x_{\theta_L,\theta_L}, \theta_L)} = 0 \Rightarrow \frac{1}{2}\mu_2(1 - (1 - p(\mu_1))^2) - \varphi_1^2 = 0 \tag{64}$$

$$\frac{\partial \mathcal{L}}{\partial W^2(x_{\theta_L,\varnothing}, \varnothing)} = 0 \Rightarrow \frac{1}{2}\mu_2(1 - p(\mu_1))^2 - \varphi_2^2 + \varphi_3^2 = 0 \tag{65}$$

$$\frac{\partial \mathcal{L}}{\partial W^1(x_{\theta_H,\varnothing}, \varnothing)} = 0 \Rightarrow \frac{1}{2}\mu_2(1 - p(\mu_1))^2 - \varphi_3^2 + \Pi_3 + \Pi_4 = 0 \tag{66}$$

$$\frac{\partial \mathcal{L}}{\partial W^2(x_{\theta_H,\theta_H}, \theta_H)} = 0 \Rightarrow \frac{1}{2}\mu_2(1 - (1 - p(\mu_1))^2) - \varphi_4^2 - \Pi_3 - \Pi_4 = 0 \tag{67}$$

$$\frac{\partial \mathcal{L}}{\partial S^1(x_{\theta_L,\theta_L},\theta_L)} = 0 \Rightarrow -(\mu_1-\mu_2)p(\mu_1)+\lambda_1^1+\alpha(\mu_1-\mu_2)p(\mu_1)=0 \tag{68}$$

$$\frac{\partial \mathcal{L}}{\partial S^1(x_{\theta_L,\varnothing},\varnothing)} = 0 \Rightarrow -(\mu_1-\mu_2)(1-p(\mu_1))+\lambda_2^1+\alpha(\mu_1-\mu_2)(1-p(\mu_1))=0 \tag{69}$$

$$\frac{\partial \mathcal{L}}{\partial S^1(x_{\theta_H,\varnothing},\varnothing)} = 0 \Rightarrow -(\mu_1-\mu_2)(1-p(\mu_1))+\alpha(\mu_1-\mu_2)(1-p(\mu_1))+\lambda_3^1-\Pi_1=0 \tag{70}$$

$$\frac{\partial \mathcal{L}}{\partial S^1(x_{\theta_H,\theta_H},\theta_H)} = 0 \Rightarrow -(\mu_1-\mu_2)p(\mu_1)+\lambda_4^1+\alpha(\mu_1-\mu_2)p(\mu_1)+\Pi_1=0 \tag{71}$$

$$\frac{\partial \mathcal{L}}{\partial S^2(x_{\theta_L,\theta_L},\theta_L)} = 0 \Rightarrow -\mu_2p(\mu_1)+\lambda_1^2+\alpha\mu_2p(\mu_1)=0 \tag{72}$$

$$\frac{\partial \mathcal{L}}{\partial S^2(x_{\theta_L,\varnothing},\varnothing)} = 0 \Rightarrow -\mu_2(1-p(\mu_1))+\lambda_2^2+\alpha\mu_2(1-p(\mu_1))=0 \tag{73}$$

$$\frac{\partial \mathcal{L}}{\partial S^2(x_{\theta_H,\varnothing},\varnothing)} = 0 \Rightarrow -\mu_2(1-p(\mu_1))+\alpha\mu_2(1-p(\mu_1))+\lambda_3^2-2\Pi_3-\Pi_4-\Pi_5=0 \tag{74}$$

$$\frac{\partial \mathcal{L}}{\partial S^2(x_{\theta_H,\theta_H},\theta_H)} = 0 \Rightarrow -\mu_2p(\mu_1)+\lambda_4^2+\alpha\mu_2p(\mu_1)+2\Pi_3+\Pi_4+\Pi_5=0 \tag{75}$$

$$\frac{\partial \mathcal{L}}{\partial \delta} = 0 \Rightarrow -\mu_2p(\mu_1)(1-p(\mu_1))+\alpha\mu_2p(\mu_1)(1-p(\mu_1))+\Pi_4-\Pi_5=0 \tag{76}$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \Rightarrow \mu_2p(\mu_1)(1-p(\mu_1))+\gamma-\alpha\mu_2p(\mu_1)(1-p(\mu_1))-\lambda_3^2-\Pi_4+\Pi_5=0 \tag{77}$$

As in previous proofs, it is immediate to see that $e_{1,1}=e_{3,1}=e_{4,1}=1$ and $e_{1,2}=e_{3,2}=e_{4,2}=1$. At the same time, using a similar reasoning to previous proofs, we can see that $W^1(x_{\theta_L,\theta_L},\theta_L)=W^2(x_{\theta_L,\theta_L},\theta_L)=\frac{1}{2}$, $W^1(x_{\theta_L,\varnothing},\varnothing)=\frac{e_{2,1}^2}{2}$, and $W^2(x_{\theta_L,\varnothing},\varnothing)=\frac{e_{2,2}^2}{2}$. So the low type agents do not receive any rents under the optimal contract. From the FOCs of $W^1(x_{\theta_H,\varnothing},\varnothing)$ and $W^1(x_{\theta_H,\varnothing},\varnothing)$ we have that both (IC3) and (IC3') are binding.

$$W^1(x_{\theta_H,\varnothing},\varnothing)=\frac{1-\Delta\theta^2}{2}+e_{2,1}\Delta\theta$$

$$W^2(x_{\theta_H,\varnothing},\varnothing)=\frac{1-\Delta\theta^2}{2}+e_{2,2}\Delta\theta$$

Before continuing with the proof, notice that if $\alpha=1$, all the $\{\lambda_i^j\}_{i=1,2,3,4}^{j=1,2}$ and the $\{\Pi_k\}_{k=1,3,4,5}$, multipliers will be zero. This implies that the contract is the non-collusion contract, which by assumption violates the collusion constraints.

For this reason, assuming that any $\{\lambda_i^j\}_{i=1,2}^{j=1,2}$ is zero makes $\alpha=1$ and drives to a contradiction. Hence, $\lambda_1^1>0$, $\lambda_2^1>0$, $\lambda_1^2>0$, $\lambda_2^2>0$. This implies that the associated wages are equal to zero. Adding equations (18) and (19), we can argue that $\lambda_3^1 > 0$, and then $S^1(x_{\theta_H,\varnothing}, \varnothing) = 0$. The same logic applies to equations (22) and (23) which mean that $S^2(x_{\theta_H,\varnothing}, \varnothing) = 0$.

Adding (25) and (26), we can see that $\gamma > 0$, which implies that $\beta = 0$. Hence, substituting into the Participation Constraint of the Monitor we get,

$$(\mu_1 - \mu_2)p(\mu_1)S^1(x_{\theta_H,\theta_H}, \theta_H) + \mu_2 p(\mu_1)(S^2(x_{\theta_H,\theta_H}, \theta_H) + (1 - p(\mu_1))\delta) = 2\bar{V}$$

If $\Pi_1 = 0$, by (20), $\alpha = 1$, which is a contradiction as already stated. Hence, $\Pi_1 > 0$, and (CC1) binds. Using the same reasoning, if $\Pi_4 = \Pi_5$, equation (25) shows that $\alpha = 2$, which is a contradiction with equation (24) as it requires that some multiplier is negative. Hence, $\Pi_4 > \Pi_5 \geq 0$, and (CC4) binds.

Now assume that $\Pi_5 = 0$. This means that (CC5) is not binding and therefore (CC3) is also slack. Hence, $\Pi_3 = 0$. Adding equations (24) and (25) we can see that $\alpha = 1$ reaching again a contradiction. This implies that all the collusion constrains and the participation constraint is binding.

Assume now that $W^1(x_{\theta_H,\theta_H}, \theta_H) > \frac{1}{2}$ (a similar logic applies to $W^2(x_{\theta_H,\theta_H}, \theta_H)$). Then, $\varphi_4^1 = 0$ and, by (12), $\Pi_1 = (\mu_1 - \mu_2)p(\mu_1)$. Using a similar procedure to previous proof, we can argue that, under this assumption, $e_{2,1} = 1 - \frac{\Delta\theta}{1-p(\mu_1)}$. If both $W^1(x_{\theta_H,\theta_H}, \theta_H)$ and $W^2(x_{\theta_H,\theta_H}, \theta_H)$ are larger than $\frac{1}{2}$, the optimal contract coincides with the collusion contract, which by assumption does not bind in the participation constraint of the monitor.

As in previous proofs, assume that, $W^1(x_{\theta_H,\theta_H}, \theta_H) = \frac{1}{2}$, $W^2(x_{\theta_H,\theta_H}, \theta_H) = \frac{1}{2}$. From the collusion constraints,

$$S^1(x_{\theta_H,\theta_H}, \theta_H) = \Delta\theta\left(e_{2,1} - \frac{\Delta\theta}{2}\right)$$
$$S^2(x_{\theta_H,\theta_H}, \theta_H) = \frac{1}{2}\Delta\theta\left(e_{2,2} - \frac{\Delta\theta}{2}\right)$$

Notice that the whole maximization problem is written in terms of $e_{2,1}$ and $e_{2,2}$ with (PCM) being binding.

$$(\mu_1-\mu_2)\left(e_{2,1}-\frac{\Delta\theta}{2}\right)+\mu_2\left(1-\frac{1}{2}p(\mu_1)\right)\left(e_{2,2}-\frac{\Delta\theta}{2}\right)=\frac{2\bar{V}}{\Delta\theta p(\mu_1)}.$$

Hence, I can rewrite the whole program and maximize as a function of only these two variables with λ being the multiplier associated to (PCM). Call this new lagrangian $\tilde{\mathcal{L}}$. This means that,

$$\frac{\partial\tilde{\mathcal{L}}}{\partial e_{2,1}}=0\Rightarrow(\mu_1-\mu_2)(1-p(\mu_1)(1-e_{2,1}-\Delta\theta)+(\mu_1-\mu_2)\lambda=0$$
$$\frac{\partial\tilde{\mathcal{L}}}{\partial e_{2,1}}=0,\Rightarrow\frac{1}{2}\mu_2(1-p(\mu_1))^2(1-e_{2,2}-\Delta\theta)+\mu_2\frac{1}{2}\left(2-p(\mu_1)\right)\lambda=0.$$

This allows me to express λ as a function of $e_{2,1}$ and substitute.

$$\lambda=(1-p(\mu_1)(e_{2,1}-1+\Delta\theta)$$

$$\frac{(2-p(\mu_1))}{1-p(\mu_1)}e_{2,1}-\frac{1-\Delta\theta}{1-p(\mu_1)}=e_{2,2}$$

Substituting back into the (PCM) we get,

$$e_{2,1}=\frac{\frac{2\bar{V}}{\Delta\theta p(\mu_1)}+(\mu_1-\mu_2)\cdot\frac{\Delta\theta}{2}+\mu_2\left(1-\frac{p(\mu_1)}{2}\right)\left(\frac{1-\Delta\theta}{1-p(\mu_1)}+\frac{\Delta\theta}{2}\right)}{(\mu_1-\mu_2)+\mu_2\left(1-\frac{p(\mu_1)}{2}\right)\cdot\frac{2-p(\mu_1)}{1-p(\mu_1)},}$$

and,

$$e_{2,2}=\frac{2-p}{1-p}\cdot\frac{\frac{2\bar{V}}{\Delta\theta p}+(\mu_1-\mu_2)\cdot\frac{\Delta\theta}{2}+\mu_2\left(1-\frac{p}{2}\right)\left(\frac{1-\Delta\theta}{1-p}+\frac{\Delta\theta}{2}\right)}{(\mu_1-\mu_2)+\mu_2\left(1-\frac{p}{2}\right)\cdot\frac{2-p}{1-p}}-\frac{1-\Delta\theta}{1-p}.$$

□