

Managing Loyalty in Hierarchical Organizations ^{*}

Hanzhe Li[†] Jin Li[‡] Arijit Mukherjee[§] Luis Vasconcelos[¶]

April 28, 2026

Abstract

Conventional wisdom regards employee loyalty to leadership as a virtue, but in hierarchical organizations it is a double-edged sword. Excessive loyalty of middle management towards the top weakens frontline workers' incentives. We show how loyalty can be managed through well-designed communication and decision-making protocols. In the optimal contract, either the middle manager is induced to align more with the bottom, or the top may directly intervene to motivate the workers. We highlight how the middle manager's loyalty to organization interacts with his loyalty to top management, and how his personal characteristics affect the management style within the organization.

JEL Classification Numbers: D82, M5

Keywords: Middle manager, relational contract, strategic communication

^{*}An earlier version of the manuscript was circulated under the title "Real and Revealed Loyalties in Organizations." For their helpful comments and suggestions, we thank Ricardo Alonso, Georgy Artemov, Daniel Barron, Jean de Bettignies, Christoph Carnehl, Wouter Dessein, Mathias Fahn, Niko Matouschek, Harry Pei, Michael Powell, João Ramos, and the seminar/workshop participants at Michigan State University, University of Melbourne, University of Queensland, 2024 Economic Theory Festival (UNSW), 10th Workshop on Relational Contracts (2024), 2024 Org Econ Workshop (UNSW), 2025 Australasian Economic Theory Workshop (Sydney), SIOE conference (2025), 2025 World Congress of the Econometric Society, CEPR IMO-ESF joint workshop (2025), and UM-MSU Theory Day workshop (2025). All errors that remain are ours.

[†]The University of Hong Kong. Email: hanzhe.li@connect.hku.hk.

[‡]The University of Hong Kong. Email: jli1@hku.hk.

[§]Michigan State University. Email: arijit@msu.edu

[¶]University of Technology Sydney. Email: luis.vasconcelos@uts.edu.au

1 Introduction

Following the seminal work by Hirschman (1970), a large literature has explored the relationship between employee loyalty to the organization and organizational performance (see, e.g., Guillon and Cezanne, 2014, for a survey). However, little has been studied about the relationship between organizational performance and the distribution of loyalty within its hierarchy, where an employee's strong loyalty to one tier may come at the expense of his loyalty towards another. This issue is most pronounced at the level of middle management that serves as a critical link between the top management and the frontline staff. And for a middle manager positioned "between [an organization's] operating core and the apex" of the establishment (Mintzberg, 1989; p. 98), a close proximity to the top leadership typically requires distancing from the bottom.

This tension has particularly sharp implications in innovative organizations. Innovation often comes from frontline employees with deep knowledge of the organization's day-to-day operations, and the headquarters relies on mid-level managers to identify and elevate promising ideas. On the other hand, employees rely on middle managers to promote their initiatives with the headquarters and ensure that adequate resources are committed to bring their projects to fruition. As developing new ideas and initiatives is costly and inherently risky, and recognition is typically earned only if the project is ultimately implemented, the employees may lack initiative for innovation unless they have a middle manager "on their side".¹ Thus, beyond their routine functions—including strategy implementation, facilitating adaptation, and coordinating across units—middle managers serve a crucial "dual role". They must evaluate projects and relay information to the top and also motivate the frontline workers to take initiatives that advance leadership's objectives.

To succeed in this role, a middle manager needs to maintain credibility across all levels of the hierarchy, but the same loyalty that makes a manager credible to senior leadership may extinguish initiative below. A manager who is closely aligned with the top may be trusted as a screener but reluctant to sponsor subordinates. On the other hand, a manager who is too closely aligned with frontline workers may preserve initiative below yet cease to be informative to the top. In this paper we study how such distribution of loyalty within the organization affects organizational performance, how loyalty can be managed to enhance efficiency, and how loyalty within the organization interacts with loyalty towards the organization.

We begin by exploring how a middle manager's alignment of his preferences, i.e., his "loyalty", with those of the top management, along with his competence in screening good projects, i.e., his "skill", affect his ability to navigate his dual role and influence firm performance.

We develop a simple static model of a hierarchical organization consisting of a principal, a

¹For example, Bensaou (2021) documents how the middle managers at IBM act as "powerful innovation cheerleaders and coaches", propelling the organization's innovative success. Similarly, Intel's pivot from memory chips to microprocessors in the 1980s was driven by middle managers who identified new opportunities and influenced strategic decisions (Burgelman, 1994; Foss and Klein, 2022). Conversely, in the 1990s and 2000s innovation declined at top U.S. information technology firms like Microsoft, Apple, and Xerox partly because companies began valuing salespeople over product engineers, reducing the engineers' motivation to innovate. As Steve Jobs, the co-founder of Apple Inc., notes: "[w]hen the sales guys run the company, the product guys don't matter so much, and a lot of them just turn off" (Isaacson, 2011; p. 569).

(middle) manager, and an agent (or “junior worker”) to analyze this issue. The agent exerts effort to generate a project that may yield certain benefits to all players. The agent obtains a benefit whenever the project is implemented, but the project’s value to the manager and the principal is a priori unknown to all parties and is revealed only after its implementation. The manager privately obtains a (noisy) signal on the project’s value to the principal and recommends whether to continue or cancel it. Upon receiving the manager’s recommendation, the principal decides whether to proceed with the project. The project’s value to the principal and the manager may not be perfectly aligned; so, they may have conflicting interests regarding the continuation of the project. The manager’s “loyalty” to the top is reflected by the alignment of the project values. The manager’s “skill” is captured by the precision of his signal. In the spirit of the delegation literature, we assume that the manager’s report is “cheap talk” and that monetary incentive contracts are infeasible. This allows us to bring to the fore the importance of the middle managers’ dual loyalty for firm performance and how to manage it through organizational design.

The static model delivers a key insight: the middle manager is effective in performing his dual role only if his loyalty to the top lies within a moderate range. A middle manager who is too loyal to the workers, becomes undiscerning in his reporting and loses credibility with the principal. Conversely, a middle manager who is too loyal to the top becomes overly selective in his recommendations. In both cases, the manager fails to motivate the agent to take initiative as the agent correctly anticipates that the projects he generates have a low chance of getting implemented. But a moderately aligned manager can balance credibility with the top and motivation for the agent, enabling him to link the two tiers of the organization and improve its performance. Furthermore, the range of loyalty for which the manager remains effective expands with his skill. Thus, a middle manager who is known to be “very loyal to the top” can still motivate effort below if he is sufficiently competent. In this sense, loyalty and skill complement each other.²

The above observation raises the following question: when a manager’s excessive loyalty (either to the top or to the bottom) renders him ineffective, how can loyalty be managed through organizational design to restore performance? The case in which the manager is too closely aligned with the bottom is conceptually similar to the delegation problem studied in principal–agent settings (see, e.g., Gibbons et al. 2012; Li et al. 2017), with the distinction that it arises here within a hierarchical organizational structure. The case more specific to hierarchical organizations is excessive loyalty to the top, as closer alignment with the top often entails greater distance from the bottom.³ We therefore focus on this case in the main analysis and return to the former later in the paper.

In particular, we examine which communication and project implementation protocols firms

²This observation stands in contrast to a class of political economy models in which a leader faces a trade-off between appointing a more competent and a more loyal agent (Egorov and Sonin, 2011; Zakharov, 2016; Dessein and Garicano, 2025). In these models, loyalty and competence are treated as substitutes, so that an increase in one dimension calls for a compromise on the other.

³Such excessive loyalty may reflect, for example, the manager’s career ambitions to climb the ranks of the organization’s hierarchy. Motivating innovation may require supporting riskier projects with a higher chance of failure. A careerist middle manager may be reluctant to support such projects as a record of past failure may hinder his aspirations for future promotions.

can adopt to moderate managerial loyalty and foster worker initiative. We explore this issue in the context of a long-term relationship between the principal, the manager, and the agent, where more nuanced protocols may be sustained. The long-term relationship is modeled as a relational contract between the three players, where they play the aforementioned static game in each period.

Our analysis yields two salient results. First, the manager’s loyalty can be tempered in two natural ways. In the optimal contract, either the manager is induced to behave as if he were less loyal to the top than he really is or the principal is induced to behave in a way that offsets the manager’s excessive loyalty. In the former case, the manager softens his recommendation by recommending more projects, and the principal follows his advice. We call this a “mediation contract” as the manager adjusts his behavior to mediate between the agent and the principal. In the latter case, the principal intervenes by directly implementing some projects (without considering the manager’s report) but otherwise adheres to the manager’s recommendation. We call this an “intervention contract.”

These two contracts mirror real-world management strategies that seek to foster innovation. In line with mediation contracts, firms often promote a culture of risk-taking to encourage idea generation. For example, Farson and Keyes (2002) document how CEOs at GE, Coca-Cola, and Lockheed created “risk-friendly workplace” where the managers were encouraged to pursue ideas that may have a high risk of failing. Similarly, House and Price (2009) describe how HP division managers were informally expected to allocate part of their R&D budgets to high-risk exploratory projects.⁴ In line with intervention contracts, direct interventions by leadership are also commonplace. Thomke (2002) details how Bank of America’s headquarters allocated funds to selected branches to develop new products and services, explicitly encouraging high failure rates to spur experimentation. In a study of a fast-growing social-media marketing firm, Turco (2016) highlights “Hack Night” events where employees pitched ideas directly to top executives, often receiving immediate support to pursue them. The corporate policy of granting discretionary time to employees to pursue any project they perceive as beneficial to the organization, such as Google’s “20% rule” used in the early 2000s, also echoes the structure of intervention contracts.

Second, we show that the optimality of each of these two types of contract depends on the characteristics of the middle manager. In this setup, a key parameter is the value a player obtains from a successful project relative to his/her outside option. In particular, if the ratio of project value to outside option for the manager is sufficiently small, the mediation contract is not feasible. In contrast, if this ratio is large enough, the mediation contract becomes (weakly) optimal. Otherwise, either mediation or intervention (or both) may be optimal depending on the manager’s loyalty and skill. When the manager’s skill and loyalty to the top are both sufficiently high, only intervention is optimal; when both are sufficiently low, only mediation is optimal; and finally, if the manager is too loyal to the top but incompetent, then no relational contract (that induces the agent to exert effort in project generation) can be sustained.

⁴See Manso (2011) for a theoretical model of incentives for innovation that highlights the optimality of tolerance (and even reward) for failure.

Our results offer two important insights. First, the optimal management of loyalty *within* the organization depends on the middle manager's loyalty *to* the organization. A manager can effectively mediate between the workers and top management only if he exhibits sufficient organizational loyalty, as reflected by his project value to outside option ratio. In other words, only the middle managers who place a high enough value on working for the organization may be induced to adjust their behavior to motivate the frontline workers. Moreover, in the optimal contract, the manager's loyalty to the organization interacts with his loyalty to the top management. A middle manager perceived as too loyal to the top can still mediate effectively provided he also has strong loyalty to the organization itself.

Second, in the optimal contract, the characteristics of the middle manager, i.e., his loyalty and skill, shape the nature of communication and decision-making within the organization. Under mediation, the principal always follows the manager's recommendation, and the manager slants his recommendation in the agent's favor to sharpen incentives. That is, the principal effectively delegates the task of incentivizing the agent to the middle manager. But under intervention, the principal takes a direct role in incentivizing the junior workers by occasionally deciding on their projects without considering the manager's recommendation. Thus, our result indicates that the "managerial style", e.g., the pattern of decision-making, that one may observe in an organization may stem from the personal characteristics of the middle management rather than that of the CEO herself. This finding is in sharp contrast to the literature on managerial fixed effects that explores correlations between the corporate policies a CEO adopts and her own characteristics (Bertrand and Schoar, 2003; Dessein and Santos, 2021; Lo et al., 2022).

While our model centers on middle managers in hierarchical organizations, the problem of dual loyalty arises in a broader set of institutional contexts. In industrial relations, for instance, the union steward (or "griever") acts as the primary advocate for the workforce, channeling employee grievances to management. Stewards frequently possess private information regarding the strength of individual cases and exercise discretion over the intensity with which grievances are pursued. Effectiveness in this role hinges on maintaining credibility with the workers' union, i.e., being perceived as aligned with workers' interests rather than as agents of management, while simultaneously preserving a constructive working relationship with management. As Dalton (1959) observes in his classic ethnographical work on workplace politics and managerial relationships, "fiery individuals" who "were bitter towards management" rarely became effective grievors, and successful grievors "held as a principle that political subtlety was necessary for success in office... You cannot have industrial relations without giving and taking on both sides" (p. 116).

A closely related tension arises in the supervision of "street-level bureaucrats" (Lipsky, 2010), public service workers such as police officers, teachers, and social workers who interact directly with community members. In this setting, supervisors function as the relevant middle managers, intermediating between the relevant public agency's leadership and the frontline agents. As emphasized in the literature, while frontline workers are tasked with implementing rules set by agencies, they retain substantial discretion to adapt policies to local conditions. An effective supervi-

sor must balance competing allegiances. Strict enforcement of agency’s directives may constrain frontline discretion and undermine effective adaptation to local conditions. Conversely, excessive alignment with frontline workers may compromise the consistency and reliability of policy implementation.⁵

Across these settings, effective performance requires that the intermediary avoid extreme alignment with either side of the hierarchy and maintain credibility with both constituencies. Our analysis formalizes this tension and shows that such moderation can be achieved either through the selection of agents predisposed to moderate alignment or through the design of incentives that tempers loyalty and induce moderation. And such designs, in turn, may shape the nature of communication and decision-making within the organization and its overall performance.

Related literature: The extant literature on loyalty, as mentioned earlier, focuses on the principal-agent relationships where loyalty is sought by the principal and shown by the agent. Thus, it has little to say about how the alignment of loyalty within a hierarchical organization affects the organizational performance—a particularly salient question in the context of middle managers whose loyalty to the top management typically comes at the expense of the rank-and-file staff, and vice versa. Our paper attempts to fill this gap. Specifically, we examine how a middle manager’s loyalty alignment *within* and *to* the organization drive their effectiveness, and how loyalty could be managed to enhance organizational performance.

The importance of middle managers for firm performance has been long recognized in the management literature (Dalton, 1959; Chandler, 1977; Burgelman, 1983a, 1983b, 1991, 1994; Mintzberg, 1989; Osterman, 2009; Jaser, 2021) and, more recently, in economics (Lazear et al. 2015; Bianco and Giorcelli, 2022; Fenizia, 2022; Friebe et al. 2022; Friebe and Raith, 2022; Metcalfe et al. 2023; also see Roberts and Shaw, 2022, for a survey of this literature). Recent empirical studies have shown that middle managers can reduce turnover with better interpersonal skills (Hoffman and Tadelis, 2021) and improve productivity via attention to worker-to-task matches (Adhvaryu et al. 2022). Additionally, the managers’ tenure, cognitive skills, and autonomy positively affect their rate of learning on the job and enhance production efficiency (Adhvaryu et al. 2023). However, the critical role of the middle managers in linking an organization’s top management with its frontline workers has not received much attention in the literature. In particular, the literature offers scant insight into the personal characteristics that make the middle managers succeed in such a role that demands “dual loyalty.”

Our paper relates to several strands of literature in organizational and personnel economics. Various authors have highlighted how incentives to innovate are stronger when the agents are more assured of support from their supervisors (Rotemberg and Saloner, 1993, 2000; Aghion et al. 2013; Nguyen, 2023). In this strand of papers, our setup is closest to the two papers by Rotemberg and Saloner. Rotemberg and Saloner (1993) consider a model of hierarchical organization (with

⁵This observation is reminiscent of Canice Prendergast (2002, 2003), who shows that when bureaucrats are monitored through customer complaints, it may be optimal to limit feedback in order to prevent them from catering to the customers’ demands irrespective of their merit.

shareholders, CEO, and a manager) to highlight the value of a CEO who is altruistic towards the manager, who must invest in developing new projects. However, in their model, the top rung of the organization, i.e., the shareholders, does not directly participate in the production process. Hence, the CEO's job does not involve linking the top and the bottom rungs of the organization and does not require maintaining dual credibility. Rotemberg and Saloner (2000) show that a CEO with a known bias towards certain types of projects may be more effective in motivating the agents, as projects that align with the CEO's bias have a greater chance of being implemented. In their model, all rungs of the organization participate in the production process. In contrast to our paper, they abstract away from the dual role of middle management and highlight the value of "unbiased" middle managers who may allocate funds to projects that may not align with the CEO's vision but nonetheless could become profitable for the firm.⁶

There is a vast literature on incentives in hierarchical organizations that has primarily focused on issues of monitoring and delegation (for a survey, see Mookherjee, 2013). On the topic of monitoring, this literature explores the role of middle managers as supervisors to the employees, where the supervisory task itself may be subject to a moral hazard problem. Following a seminal contribution by Tirole (1986), several papers have explored the implications of such contracting frictions on the provision of incentives within the organization (Kofman and Lawarree, 1993; Olsen and Torsvik, 1998; Burguet and Che, 2004). Bearing resemblance to our modeling framework, a more recent contribution by Troya-Martinez and Wren-Lewis (2023) models the firm-manager-supplier relationship as a relational contract, where the manager and the supplier may collude on the price and quality of the product. Allowing for monetary transfers, they argue that side payments or "kickbacks" from the agent to the manager can facilitate incentive provision by decreasing the manager's temptation to renege on her promise. The literature on delegation, on the other hand, explores the trade-off between loss of incentives/coordination and loss of control as the firm decides whether to delegate decision rights to a biased but potentially more informed manager (Melumad et al. 1995; Aghion and Tirole, 1997; Baker et al. 1999; Athey and Roberts, 2001; Alonso and Matouschek, 2007a, 2007b; Alonso et al., 2008; Dessein, 2002; Rantakari, 2008, 2012; Li et al., 2017; Lipnowski and Ramos, 2020).

Our paper contrasts with and contributes to this literature by highlighting how the effectiveness of a middle manager requires careful management of his dual loyalty, but we abstract away from the issue of monitoring the agents' effort and the control-coordination trade-off with decentralized decision-making.

This article also contributes to the literature on strategic information revelation in relational contracts. Studying a varied range of environments, several authors have shown how a deliberate manipulation of information can make relational contracts more efficient (see. e.g., Fuchs, 2007; Kvaloy and Olsen, 2009; Fong and Li, 2017; Orlov, 2022; Wong, 2023; and, for an application in dynamic incentive provisions, Vong, 2025). When messages are cheap talk, repeated interaction

⁶In contrast to most of this literature, including our paper, that treats loyalty as exogenous characteristics or bias, Board (2011) shows that in repeated buyer-seller relationships, loyalty may arise endogenously in response to hold-up risks, leading the buyer to favor certain group of suppliers over the others.

may also discipline the sender to communicate truthfully (Kolotilin and Li, 2021; Kuvalekar, et al. 2022). However, these authors study principal-agent relationships rather than communication in organizational hierarchies, and we explore how strategic communication between the middle manager and the top management affects the incentives for the bottom worker.

In the aforementioned literature, Troya-Martinez and Wren-Lewis (2023) and Fong and Li (2017) deserve particular mention as, to the best of our knowledge, these are the only two papers that consider relational contracts in an organizational hierarchy. However, in both papers, at least one player does not have any dynamic incentives as they do not actively participate in the relationship. In Troya-Martinez and Wren-Lewis (2023), the principal chooses some parameters of the contract at the beginning of the game and lets the middle manager and the agent engage in a relational contract, whereas in Fong and Li (2017), the middle manager is modeled as a non-strategic “supervisor” who simply relays information as per a pre-specified reporting rule. In contrast, we consider a setup where the players in all three rungs of the organization actively participate in a relational contract, and explore how the interplay among their dynamic incentive constraints shapes the effectiveness of the middle manager and influences communication and decision-making within the organization.

The remainder of the article is organized as follows: In [Section 2](#), we present a static model that serves as a baseline for our analysis, and it is analyzed in [Section 3](#). [Section 4](#) embeds the baseline model in an infinitely repeated game, and [Section 5](#) explores the feasibility of mediation and intervention contracts. The optimal relational contract is characterized in [Section 6](#). A final section, [Section 7](#), presents a few extensions of our analysis and draws a conclusion. All proofs are given in the Appendix.

2 A baseline model

In this section, we develop a simple model of a hierarchical organization where a principal $\mathcal{P}$ (or “firm”) hires a (middle) manager $\mathcal{M}$ and an agent $\mathcal{A}$ (or “junior worker”) who may work together on a project. The model illustrates how the middle manager’s “loyalty” and “skill” affect his credibility with his supervisor and the work incentives of his subordinates, and it will later serve as the “stage game” in a more general model with repeated interaction.

We elaborate on the model by describing its three key components: *Technology*, *information structure and reporting*, and *payoffs*.

TECHNOLOGY: The agent (publicly) exerts costly effort $e \in \{0, 1\}$ to generate a project that may yield private value to all three players. The cost of effort is $c(e)$, where $c(0) = 0$ and $c(1) = c$. A project is generated if and only if $e = 1$. Once a project is generated, the manager evaluates its potential value and reports to the principal on whether she should implement it (we will elaborate on the manager’s reporting decision shortly). If the project is implemented, the agent obtains a value of b . The project’s value to the principal, $V_{\mathcal{P}} \in \{0, Y\}$, is stochastic where $V_{\mathcal{P}} = 0$ or Y with equal probability. However, the project’s value to the manager, $V_{\mathcal{M}} \in \{0, y\}$, depends on the project’s

type $\omega \in \{C, D\}$. If $\omega = C$, the manager's value is "congruent" with that of the principal: $V_M = y$ if and only if $V_P = Y$. If, instead, $\omega = D$, their values "diverge" as $V_M = y$ regardless of V_P . The state ω is a priori unknown to all parties who hold a common prior $\Pr(\omega = C) = \alpha$.

Note that for high α the manager is more likely to benefit from the project only if the principal also benefits from it. In that case, the manager would be more discerning in his recommendation regarding the project's implementation. In contrast, for low α , the manager is more likely to always benefit from the project's implementation—as is the agent—and would be indiscriminate with his recommendation to implement the project. Thus, α reflects how aligned the manager's preference is with that of the principal vis-à-vis the agent, and one may interpret α as the "loyalty" of the manager towards the principal (we will further elaborate on this issue in the conclusion of this section).

INFORMATION STRUCTURE AND REPORTING: Once a project is generated, the manager observes its type ω and a signal $s \in \{0, 1\}$ on the principal's payoff from the project, where

$$\Pr(s = 1 \mid V_P = Y) = \sigma \text{ and } \Pr(s = 1 \mid V_P = 0) = 0.$$

We interpret $\sigma \in (0, 1)$ as the "skill" of the manager as it captures the manager's ability to identify good projects for the principal. Though the signal is noisy, $s = 1$ indicates a high-value project for the principal with certainty. Upon observing the signal, the manager sends a report $r \in \{0, 1\}$ to the principal. The report is "cheap talk" that need not align with the manager's observed signal. Observing the manager's report, the principal decides whether to implement the project or cancel it. We denote the principal's implementation decision by $\iota \in \{0, 1\}$, where $\iota = 1$ if the project is implemented and 0 otherwise.

The table below summarizes the project values for the principal and the manager given the project type and the manager's signal.

	$s = 1$ (prob. $\frac{1}{2}\sigma$)	$s = 0$ (prob. $1 - \frac{1}{2}\sigma$)
$\omega = C$ (prob. α)	(Y, y)	(Y, y) or $(0, 0)$
$\omega = D$ (prob. $1 - \alpha$)	(Y, y)	(Y, y) or $(0, y)$

Table 1: Principal and manager's project values, (V_P, V_M) , given the state and signal (ω, s) .

PAYOFFS: All players are risk neutral. If any one of the three players decides not to participate, all players receive their outside options. We assume that the outside option for the agent is 0, whereas

the outside options of the manager and the principal are v_0 and π_0 , respectively. These are also the payoffs of the three players if the agent withholds effort, i.e., $e = 0$, and no project is generated.

Thus, if all players participate, given the project type ω , manager's signal s , and his report r , the payoffs of the agent, manager, and the principal are:

$$\hat{u} := \begin{cases} b - c & \text{if } e = 1 \text{ and } P \text{ implements the project} \\ -c & \text{if } e = 1 \text{ and } P \text{ does not implement the project} \\ 0 & \text{otherwise} \end{cases},$$

$$\hat{v} := \begin{cases} \mathbb{E}[V_{\mathcal{M}} \mid s, \omega] & \text{if } e = 1 \text{ and } P \text{ implements the project} \\ v_0 & \text{otherwise} \end{cases},$$

and

$$\hat{\pi} := \begin{cases} \mathbb{E}[V_{\mathcal{P}} \mid r] & \text{if } e = 1 \text{ and } P \text{ implements the project} \\ \pi_0 & \text{otherwise} \end{cases}.$$

respectively.

We impose the following restrictions on the parameters in order to focus on a more interesting modeling environment.

Assumption 1. (i) $\frac{1}{2}Y < \pi_0 < Y$; (ii) $\frac{1}{2}y < v_0 < y$; and (iii) $c < b < 2c$.

Assumption 1 (i) implies that it is optimal for the principal to implement the project when it is known to be of high value, but it is not worth implementing solely based on her prior belief about the project's valuation. In a similar spirit, Assumption 1 (ii) implies that the middle manager would always prefer to have the project implemented if he is sure to have a high value but would prefer to have it canceled if he only knows that the state is C , and hence, both high and low values are equally likely. Finally, Assumption 1 (iii) imposes bounds on the agent's benefit. The agent may be induced to exert effort only if his benefit exceeds his cost. The upper bound on the benefit rules out certain boundary cases and streamlines the analysis but does not have any substantive implications for our findings.

TIMELINE: The timing of the game is summarized below.

- *Stage 1.* The principal, the manager and the agent decide whether to engage in the production process. If at least one of the three players refuses to participate, the game ends. Otherwise, the game moves to Stage 2.
- *Stage 2.* The agent chooses his effort level e . If $e = 0$, the game ends. Otherwise, the game moves to stage 3.
- *Stage 3.* The manager observes the project type ω and signal s .
- *Stage 4.* The manager sends a report r to the principal.

- *Stage 5.* The principal decides whether to implement the project.
- *Stage 6.* All parties receive their payoffs, and the game ends.

STRATEGIES AND EQUILIBRIUM: The strategies of all three players have two components each. For each player, the first component of their respective strategies is a decision on whether to participate in the production process. The second component of the strategies differs across the players. For the agent, it specifies an *effort choice* $e \in \{0, 1\}$. For the manager, it specifies a *reporting policy* that maps his information on the state $\omega \in \{C, D\}$ and his signal $s \in \{0, 1\}$ to his report $r \in \{0, 1\}$. Finally, for the principal, it specifies an *implementation policy* that maps the manager's report to a decision on whether to implement or cancel the project. We use perfect Bayesian Equilibrium as our equilibrium concept.

We conclude this section with the following remarks on our modeling assumptions.

First, we model the middle manager's loyalty by the parameter α , which captures how closely his preferences regarding project implementation align with those of his superior and subordinates. Our production environment can be mapped into a wide range of real-world scenarios. Even though the agent is in charge of generating the project, the middle manager's personal characteristics may influence the type of project that the agent can generate as the manager can facilitate the process by ensuring task coordination, supplying complementary inputs and guidance. A manager may manipulate the process to ensure that the project is profitable to him regardless of its value to the principal. In this environment, α would simply reflect the likelihood that the manager refrains from such influence activities. Alternatively, one may consider a setup where the project can be evaluated on multiple dimensions. Suppose that the project has two aspects; though the middle manager may care about both, the principal only cares about the first one. In particular, if the second aspect is above a certain threshold, the manager wants to proceed with the project regardless of the other aspect of the project. But if the second aspect is below the threshold, then whether he wants to proceed or not depends on the first aspect. The manager's loyalty, α , may reflect the probability that a project's second aspect is below the threshold, and hence, the manager now cares about the first aspect only, as does the principal. Thus a more loyal manager is one with a higher threshold in the second aspect of a project, i.e., a manager who is unlikely to assign any special importance to the second aspect.⁷

Second, in our model, the manager is essential for production as he possesses the expertise necessary to assess project value—expertise that junior workers lack and that headquarters cannot feasibly replicate given its limited attention and evaluation capacity. Consequently, the agent cannot directly observe or report the signal of project quality, nor can headquarters directly assess

⁷As a concrete example, consider a firm where innovation is crucial. The firm employs researchers who have to take the initiative to develop new projects with very uncertain outcomes. Suppose the principal cares about the short-term profitability of the project. The middle manager, on the other hand, may care about the profitability and the value of scientific discovery. If the scientific discovery is novel enough, the middle manager wants the project to proceed (regardless of the profitability). If the scientific discovery is not quite significant, then the middle manager will look at the project's profitability and prefer to continue only if it is profitable to do so.

the project’s likelihood of success. To streamline the model, we therefore abstract from the middle manager’s direct participation in the production process, though in many settings the manager may facilitate project generation through direct effort and the provision of complementary inputs.

Finally, in line with the standard approach in the literature on delegation and intra-firm communication, we rule out the possibility of incentive provision through monetary transfers (see, e.g., Holmström, 1977, 1984; Aghion and Tirole, 1997; Dessein, 2002; Alonso and Matouschek, 2007a, 2007b; Alonso et al., 2008; Lipnowski and Ramos, 2020). This modeling choice highlights the salience of the middle managers’ dual role and underscores the importance of organizational design in managing their alignment of loyalty within the firm hierarchy. The absence of incentive pay may arise for various reasons. For instance, the project outcome may be subjective in nature (i.e., privately observed by the principal) or difficult to codify (as in the case of disruptive innovations), precluding its use in contingent contracts. Intra-firm communication may also be unverifiable in court, and contracting frictions—such as risk aversion and multitasking problems—may render high-powered incentive pay infeasible or counterproductive.

3 Analyzing the baseline model

This section characterizes the equilibrium in our baseline model, and, in the process, highlights the key economic effects that are in play. In particular, it illustrates how the middle manager’s loyalty and skill affect his ability to achieve the dual goals of maintaining credibility with the top management and encouraging work initiative among his subordinates as he strategically reports his information to the principal.

We begin our analysis with the following observation. From the principal’s perspective, the ideal outcome is one where the agent exerts effort, and the project is implemented if and only if the associated signal is favorable (i.e., $s = 1$). However, two frictions in our model prevent the principal from achieving this outcome. First, the principal cannot directly observe the signal on project quality and must rely on the manager’s report when deciding on project implementation. The manager, in turn, may manipulate the report to influence the principal’s decision. Second, the decision rule that results from this communication may lead to an ex-ante likelihood of project implementation that is insufficient to incentivize the agent to generate a project in the first place.

In any equilibrium of this model, these two frictions compromise the principal’s payoff. In particular, as the manager’s report is “cheap talk”, there always exists a “babbling equilibrium” where the manager’s report is completely uninformative (e.g., he sends the same report regardless of his information) and, in response, the principal always ignores the manager’s report. As no project gets implemented, the agent does not generate any project in the first place, and all players take their respective outside options. In what follows, we focus on the equilibrium that, given the middle manager’s characteristics, yields the highest payoff to the principal.

In any informative equilibrium (i.e., one where the manager’s recommendation influences the principal’s implementation decision), the manager would always recommend implementation if

$\omega = D$ or $(\omega, s) = (C, 1)$, as in both cases his value of the project (y) exceeds his outside option (v_0). Thus, the only informative communication that may be sustained in equilibrium is one where the principal implements the project only if the manager recommends implementation by sending $r = 1$ (say) and the manager chooses to send $r = 1$ except when $(\omega, s) = (C, 0)$.

Such a communication protocol induces the agent to exert effort if and only if $\Pr(\iota = 1 \mid e = 1)b = \Pr(r = 1 \mid e = 1)b \geq c$, i.e.,

$$1 - \alpha \left(1 - \frac{1}{2}\sigma\right) \geq \frac{c}{b}. \quad (IC_A)$$

(Trivially, $\Pr(\iota = 1 \mid e = 0) = 0$ as no project is generated in the first place.) One may readily check that the manager cannot gain from deviating from the above reporting strategy. As mentioned above, he obtains $y > v_0$ in all states where he reports $r = 1$, and when $(\omega, s) = (C, 0)$, his expected payoff from the project is $\frac{1-\sigma}{2-\sigma}y$, which, by Assumption 1 (ii), is smaller than v_0 , the payoff he gets by reporting $r = 0$ and having the project canceled.

Finally, consider the principal's incentive to follow the manager's recommendation. It is optimal for her to follow the manager's recommendation if and only if

$$\Pr(V_P = Y \mid r = 1)Y \geq \pi_0,$$

i.e.,

$$\frac{\frac{1}{2}\sigma + \frac{1}{2}(1-\sigma)(1-\alpha)}{1-\alpha(1-\frac{1}{2}\sigma)}Y \geq \pi_0, \quad (IC_P)$$

and $\Pr(V_P = Y \mid r = 0)Y < \pi_0$, i.e.,

$$\frac{1-\sigma}{2-\sigma}Y < \pi_0. \quad (IC'_P)$$

Note that (IC'_P) is always satisfied under Assumption 1 (i). Thus, the aforementioned strategies of the principal and the manager, along with $e = 1$, can be supported in equilibrium if and only if (IC_A) and (IC_P) are satisfied. The proposition below characterizes this equilibrium and the conditions under which it can be sustained.

Proposition 1. *The equilibrium that yields the largest payoff to the principal is characterized as follows. If $Y/\pi_0 < 2c/b$, then no project is generated, and the players obtain their outside options. Otherwise, there exists $\sigma_0 \in (0, 1]$ such that:*

- (i) *If $\sigma < \sigma_0$, no project is generated, and the players obtain their outside options.*
- (ii) *If $\sigma \geq \sigma_0$, there exist two cutoffs, α_0 and α_1 (depending on σ), where $0 < \alpha_0 \leq \alpha_1 < 1$, such that:*
 - (a) *If $\alpha \in [\alpha_0, \alpha_1]$, the agent generates a project (i.e., $e = 1$), the manager reports $r = 0$ when $(\omega, s) = (C, 0)$ and reports $r = 1$ otherwise, and the principal implements the project if and only if $r = 1$.*
 - (b) *If $\alpha \notin [\alpha_0, \alpha_1]$, no project is generated, and all players obtain their outside options.*

(c) The cutoff α_0 is decreasing in σ whereas α_1 is increasing in σ , with $\alpha_1 = \alpha_0$ if $\sigma = \sigma_0$.

Proposition 1 shows how the agent's incentives and communication between the manager and the principal are affected by the manager's characteristics, namely, his skill and loyalty to the principal. When the precision of the manager's signal is low ($\sigma < \sigma_0$), the principal is always better off taking her outside option, and hence, no project is generated. But when the signal is sufficiently precise ($\sigma \geq \sigma_0$), production may take place, provided that the manager's loyalty to the principal is in an intermediate range. However, when the principal's value of a successful project (relative to her outside option) is small or the agent's cost of effort (relative to his benefit) is high (i.e., $Y/\pi_0 \leq 2c/b$), no project can be generated regardless of the manager's characteristics.

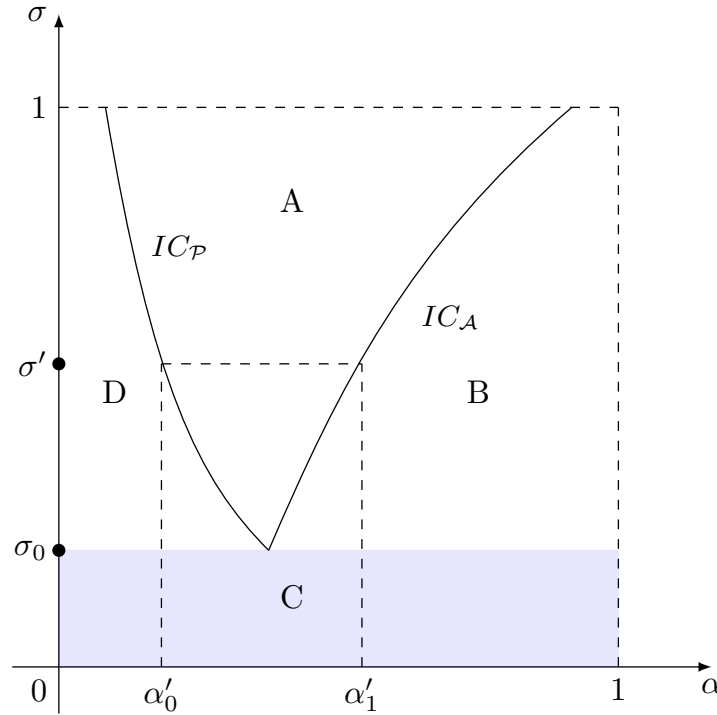


Figure 1: Equilibrium characterization in the baseline model.

Figure 1 illustrates the equilibrium characteristics (when $Y/\pi_0 > 2c/b$) as stated in the above proposition. There exists an equilibrium where the agent exerts effort if and only if the parameters (α, σ) are in region A. For example, when the manager's skill $\sigma = \sigma'$, in this equilibrium, the agent exerts effort and the principal follows the manager's recommendation, provided that the manager's loyalty to the principal, α , is in $[\alpha'_0, \alpha'_1]$. In all other regions, no equilibrium exists where the agent exerts effort to generate a project. In region B, (IC_A) is violated; in region D, (IC_P) is violated; and in region C, the signal precision is so low that the principal prefers to take her outside option rather than engage in production.

To see the intuition for this result, recall the dual goal of the manager's reporting policy: the manager needs to recommend the agent's project sufficiently often to incentivize the agent to exert

effort, but the manager may not be too indiscriminate in his recommendations lest he lose credibility with the principal. In other words, for the manager's reporting policy to be effective in meeting these two goals, both (IC_A) and (IC_P) must hold. Now, for a given "skill" level (σ), the more loyal the manager is, i.e., the higher α is, the more likely it is that the project is of value to him *only when* it is also of value to the principal. Consequently, he becomes more discerning in his recommendations; but as this lowers the likelihood that the project will be implemented, he fails to incentivize the agent. On the other hand, when the manager's loyalty to the principal is sufficiently low, he is too indiscriminate in his recommendations, and the principal disregards his report. Thus, the manager can attain the dual goal of his reporting policy in equilibrium only if he is truly in the "middle"—not too loyal to the top and also not too close to the bottom.

Interestingly, when the principal's value of a successful project (relative to her outside option) is small or the agent's cost of effort (relative to his benefit) is high, such "middle" ground ceases to exist as (IC_A) and (IC_P) cannot be met simultaneously. The minimum loyalty that the manager must have to retain credibility with the principal becomes too high, rendering him ineffective in incentivizing the agent. Thus, the agent cannot be induced to exert effort; no project is generated, and all parties take their respective outside options.

Proposition 1 also highlights a complementarity between the manager's skill and loyalty: the range of loyalty, $[\alpha_0, \alpha_1]$, that can support production in equilibrium, increases in the manager's skill, σ . In other words, both (IC_A) and (IC_P) get relaxed with higher σ . As the manager's signal becomes more precise, the principal's expected payoff from following the manager's recommendation (to implement the project) increases. The agent, on the other hand, becomes more assured that his effort will lead to project implementation, as the manager is more likely to detect a good project when one is generated. Finally, from **Proposition 1**, we can answer the following question: if the principal could recruit any middle manager by skill and loyalty, whom would she choose? Note that the principal's equilibrium payoff is increasing in both σ and α .⁸ Thus, the principal would always prefer more skilled managers over less skilled ones, and for a given skill level σ , the maximal loyalty satisfying (IC_A) , i.e., $\alpha_1(\sigma)$. Consequently, the best pick would be one with skill $\sigma = 1$ and loyalty $\alpha_1(1)$.

In what follows, we limit attention to the case where $Y/\pi_0 \geq 2c/b$ in order to focus on the more interesting modeling environment.

4 Managing loyalty

The above analysis of the baseline model highlights two types of distortions that compromise the principal's payoff: first, in region *A* the project is implemented too often relative to what the principal would prefer; second, in regions *B* and *D* the project is not implemented at all; in fact, it is not generated in the first place as the agent lacks incentives for exerting effort. The distortion in

⁸Higher σ means a good signal is more informative of high output, and higher α implies that, conditional on implementation recommendation, it is more likely that the manager has received a good signal.

regions A and D arises because the manager is not loyal enough to the principal and recommends implementation too often (in region D in particular, the manager's recommendation completely loses its credibility). In contrast, the distortion in region B stems from the fact that the manager is too loyal to the principal and does not recommend implementation often enough. We explore how such inefficiencies may be ameliorated in a long-term employment relationship where the parties may rely on relational incentives to influence the likelihood of project implementation.

The remedy for the first type of problem (regions A and D) calls for discouraging the manager from recommending projects that are beneficial to him but detrimental to the principal. Such an incentive problem is reminiscent of the seminal work by Aghion and Tirole (1997) on delegation and the literature that followed.⁹ In contrast, the remedy for the latter type of problem (region B) calls for incentivizing the manager and the principal to be more lenient in project implementation so as to encourage the agent to exert effort and generate a project. This is a novel problem that has hitherto received little attention in the literature, and it is particularly salient in the context of middle managers in hierarchical organizations where heightened loyalty to the top inherently requires distancing from the bottom. In what follows, we primarily focus on this case and address the question of how excessive loyalty of middle management to the top may be optimally modulated to restore the incentives of the frontline workers.

4.1 A model of relational contract

We consider an infinitely repeated game where the principal, manager, and agent enter into a long-term employment relationship. Time is assumed to be discrete and denoted as $t \in \{1, 2, \dots, \infty\}$. All players have a common discount factor $\delta \in (0, 1)$. In each period, the players play a stage game that is identical to the game described in our baseline model, except for the following modifications.

We assume that the project type is i.i.d across periods. In order to allow for more nuanced reporting and implementation policies, we introduce two randomization devices that the manager and the principal may use in their respective reporting and implementation decisions. After the agent exerts effort, all players observe the realization x_t of a public randomization device. In addition, the manager and the agent also observe the realization z_t of a randomization device that is not observed by the principal. Both x_t and z_t are independently and uniformly distributed on $[0, 1]$.

The manager, given his information on the state and signal (ω_t, s_t) and the realization of the randomization devices, (x_t, z_t) , sends a report $r_t \in \{0, 1\}$ to the principal. The report is observed by all the players. Given the manager's report and the realization of the public randomization device x_t , the principal makes her decision on implementation, $\iota_t \in \{0, 1\}$ ($\iota_t = 1$ if the project is implemented and 0 otherwise).

The timeline of the stage game is illustrated in [Figure 2](#). We denote the participation decision of player $i \in \{\mathcal{A}, \mathcal{M}, \mathcal{P}\}$ as $d_t^i \in \{0, 1\}$, where $d_t^i = 1$ if and only if player i decides to participate in

⁹See, e.g., Gibbons et al. (2012) for a survey of this literature.

period t . Furthermore, we assume that the project quality—that is, the principal’s realized payoff from an implemented project—is revealed at the end of the stage game for each period.

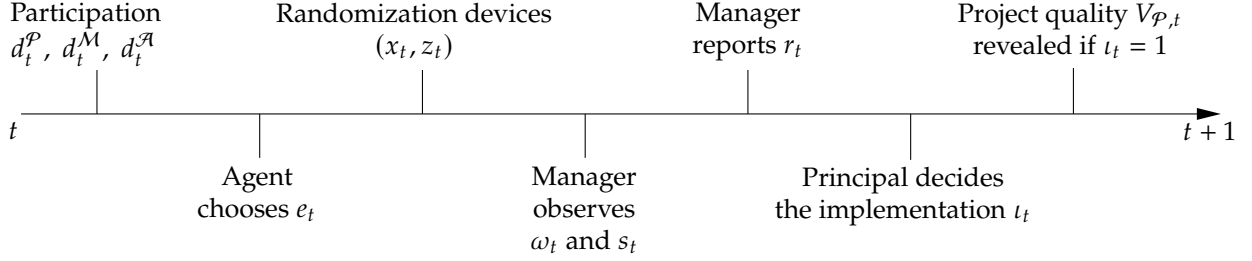


Figure 2: The timeline of the stage game in the repeated version.

HISTORIES: The particular information structure of the game leads us to define three different histories. First, let $h_t^P = \{d_t^A, d_t^M, d_t^P, e_t, x_t, r_t, l_t, V_{P,t}\}$ be the public history of period t where $V_{P,t}$ denotes the principal’s realized payoff in period t (set $V_{P,t} = \emptyset$ if $l_t = 0$), $h^{P,t} = (h_1^P, h_2^P, \dots, h_{t-1}^P)$ be the public history up to date t , and $H^{P,t}$ be the set of all possible date t histories (set $H^{P,1} = \emptyset$). Note that the set of the principal’s date t histories is $H^{P,t}$. Next, to obtain the sets of the agent’s and the manager’s date t histories, we combine their private information with the public history. Let $h_t^A := h_t^P \cup \{z_t\}$ and $h_t^M := h_t^P \cup \{z_t, \omega_t, s_t\}$ be the private history of period t observed by the agent and the manager, respectively. Denote their respective private histories up to date t as $h^{A,t} = (h_1^A, \dots, h_{t-1}^A)$ and $h^{M,t} = (h_1^M, \dots, h_{t-1}^M)$, and let $H^{A,t}$ and $H^{M,t}$ be the set of all possible private histories up to date t for the agent and the manager.

STRATEGIES: A strategy of the agent specifies a sequence of functions $\sigma^A := \{(\sigma_{t,1}^A, \sigma_{t,2}^A)\}_{t=1}^\infty$. For date t , the function $\sigma_{t,1}^A : H^{A,t} \rightarrow \{0, 1\}$ specifies his participation decision, and the function $\sigma_{t,2}^A : H^{A,t} \rightarrow \{0, 1\}$ specifies the agent’s effort choice.

Similarly, a strategy of the manager stipulates a sequence of functions $\sigma^M := \{(\sigma_{t,1}^M, \sigma_{t,2}^M)\}_{t=1}^\infty$. The function $\sigma_{t,1}^M : H^{M,t} \rightarrow \{0, 1\}$ specifies the manager’s participation decision for date t based on his history. The function $\sigma_{t,2}^M : H^{M,t} \times \{0, 1\} \times [0, 1]^2 \times \{C, D\} \times \{0, 1\} \rightarrow \{0, 1\}$ specifies the manager’s reporting decision in date t given his history, the agent’s effort e_t , the realization of the two randomizing devices (x_t, z_t) , and his observation on the project’s state and the signal (ω_t, s_t) .

Finally, a strategy of the principal stipulates a sequence of functions $\sigma^P := \{(\sigma_{t,1}^P, \sigma_{t,2}^P)\}_{t=1}^\infty$. The function $\sigma_{t,1}^P : H^{P,t} \rightarrow \{0, 1\}$ specifies the principal’s participation decision in period t given her history. The function $\sigma_{t,2}^P : H^{P,t} \times \{0, 1\} \times [0, 1] \times \{0, 1\} \rightarrow \{0, 1\}$ specifies her implementation decision for date t given her history, the agent’s effort e_t , the realization of the public randomizing device x_t , and the manager’s recommendation r_t .

PAYOFFS: With a slight abuse of notation, we denote a profile of strategies $(\sigma^A, \sigma^M, \sigma^P)$ as σ . Given a strategy profile σ and a history up to date t , $h^t \in H^{A,t} \cup H^{M,t} \cup H^{P,t}$, the expected payoffs of the

players from the continuation game starting in period t are defined as follows. For the agent:

$$u_t(h^t, \sigma) = (1 - \delta) \mathbb{E} \left[\sum_{\tau=t}^{\infty} \delta^{\tau-t} d_{\tau} \hat{u}_{\tau} \mid h^t, \sigma \right],$$

for the manager:

$$v_t(h^t, \sigma) = (1 - \delta) \mathbb{E} \left[\sum_{\tau=t}^{\infty} \delta^{\tau-t} (d_{\tau} \hat{v}_{\tau} + (1 - d_{\tau}) v_0) \mid h^t, \sigma \right],$$

and for the principal:

$$\pi_t(h^t, \sigma) = (1 - \delta) \mathbb{E} \left[\sum_{\tau=t}^{\infty} \delta^{\tau-t} (d_{\tau} \hat{\pi}_{\tau} + (1 - d_{\tau}) \pi_0) \mid h^t, \sigma \right],$$

where $d_{\tau} := d_{\tau}^{\mathcal{A}} d_{\tau}^{\mathcal{M}} d_{\tau}^{\mathcal{P}} = 1$ if all players decide to participate in production in period τ , and $d_{\tau} = 0$ otherwise. When no confusion arises, we omit the strategy profile σ in the payoff expressions. Thus, the tuple (u_1, v_1, π_1) represents the players' payoffs at the beginning of the repeated game.

We refer to $u_t(h^{\mathcal{A},t}), v_t(h^{\mathcal{M},t}), \pi_t(h^{\mathcal{P},t})$ —i.e., the agent, the manager, and the principal's expected payoffs in the continuation game conditional on the history that they individually observe—as the players' respective continuation payoffs at the beginning of period t .

EQUILIBRIUM: A pure-strategy perfect Bayesian equilibrium (referred to below simply as an "equilibrium") consists of a strategy profile $\sigma = (\sigma^{\mathcal{A}}, \sigma^{\mathcal{M}}, \sigma^{\mathcal{P}})$ and the players' posterior beliefs $\mu = (\mu^{\mathcal{A}}, \mu^{\mathcal{M}}, \mu^{\mathcal{P}})$ that stipulate each player i 's belief about the others' private histories given his/her own private history $h^{i,t}$, ($i \in \{\mathcal{A}, \mathcal{M}, \mathcal{P}\}$). In our model, the manager's private history contains more information than the agent's, and the agent's private history contains more information than that of the principal's (which is the public history of the game). Thus, the agent and the manager's beliefs about the history observed by the principal (i.e., $\mu^{\mathcal{A}}(h^{\mathcal{P},t} \mid h^{\mathcal{A},t})$, and $\mu^{\mathcal{M}}(h^{\mathcal{P},t} \mid h^{\mathcal{M},t})$) are necessarily degenerate, and so is the manager's belief about the agent's private history (i.e., $\mu^{\mathcal{M}}(h^{\mathcal{A},t} \mid h^{\mathcal{M},t})$). The posterior beliefs $\mu^{\mathcal{A}}(h^{\mathcal{M},t} \mid h^{\mathcal{A},t})$, $\mu^{\mathcal{P}}(h^{\mathcal{A},t} \mid h^{\mathcal{P},t})$, and $\mu^{\mathcal{P}}(h^{\mathcal{M},t} \mid h^{\mathcal{P},t})$ stipulate how the agent and the principal assign (non-degenerate) probabilities to the private histories of the other players, conditional on their own observed histories.

The strategies and posterior beliefs constitute an equilibrium if they satisfy the following two conditions:

- (i) **Best Response:** For each player $i \in \{\mathcal{A}, \mathcal{M}, \mathcal{P}\}$, σ^i is a best response to the other players' strategies after every private history $h^{i,t}$, given the beliefs μ .
- (ii) **Belief Consistency:** The beliefs μ are consistent with the strategy profile σ and are updated using Bayes' rule whenever possible.

We define a relational contract as an equilibrium of this repeated game. An optimal relational contract is an equilibrium that maximizes the principal's payoff at the beginning of the game, i.e.,

π_1 . As we are interested in the principal's optimal equilibrium, without loss of generality, we focus on the class of equilibria in which any observable deviation triggers the players to terminate the relationship, resulting in each player receiving their respective outside option payoffs.

4.2 Preliminary observations

As mentioned at the beginning of this section, we focus our analysis on the scenario where, in the baseline model, the agent has insufficient incentives to work because the manager is excessively aligned with the principal and does not recommend the project often enough (i.e., the region B in Figure 1). With a slight abuse of notation, we also denote this set of parameters as $B := \{(\alpha, \sigma) \mid \sigma \geq \sigma_0, \text{ and } (IC_A) \text{ is violated}\}$.

To solve the model, we begin by considering a class of contracts where, in each period, the agent exerts effort ($e_t = 1$), the project is implemented with probability 1 if the state and the signal realization $(\omega, s) \neq (C, 0)$, but when $(\omega, s) = (C, 0)$, the project is implemented with probability η . Any such contract, if feasible, must satisfy the agent's incentive constraint: $\Pr(\iota_t = 1 \mid e_t = 1)b \geq c$, i.e.,

$$1 - \alpha \left(1 - \frac{1}{2}\sigma\right) (1 - \eta) \geq \frac{c}{b}. \quad (IC_A^*)$$

Let η^* be the value of η for which (IC_A^*) binds. (As the left-hand side of (IC_A^*) increases in η , (IC_A^*) holds for all $\eta \geq \eta^*$.) Fix $\eta = \eta^*$, and denote the agent's, the manager's, and the principal's payoff associated with such a contract as u^* , v^* , and π^* , respectively. As the agent is indifferent between exerting effort and shirking, $u^* = 0$.

Lemma 1. *Suppose $(\alpha, \sigma) \in B$. Then, the following holds:*

- (i) $\pi^* \geq \pi_0$.
- (ii) When $v^* < v_0$, the only equilibrium is one in which all players take their outside options in all periods.
- (iii) When $v^* \geq v_0$, in any equilibrium, the continuation payoffs of the manager and the principal at the beginning of each period are bounded above by v^* and π^* , respectively.

Lemma 1 shows that for any equilibrium of the game where the agent exerts effort, v^* and π^* are the upper bounds on the manager's and the principal's continuation payoff at each period. In particular, π^* is the maximal equilibrium payoff for the principal. Thus, any contract in the aforementioned class (with $\eta = \eta^*$) is indeed optimal if it is self-enforcing.

Note that any contract in this class requires the implementation of the project when $(\omega, s) = (C, 0)$ with some probability. However, when $(\omega, s) = (C, 0)$, the project's expected values for both the manager and the principal are less than their outside options, and both would prefer to cancel it. Thus, a contract where such projects are implemented requires that either the manager or the principal sometimes take an action against their short-term interests so as to sharpen the agent's incentives to generate new projects.

In light of the above observation, one may consider two natural forms of relational contracts. First, it is the manager who takes an action against his short-term interests: in any given period, he recommends the implementation of the project when $(\omega, s) = (C, 0)$ with probability η^* , and the principal always adheres to the manager's recommendation. Second, it is the principal who takes such an action: in any given period, with probability η^* , she completely disregards the manager's report and implements the project generated by the agent, but otherwise (with probability $1 - \eta^*$) she adheres to the manager's recommendation (which is identical to that in the static game). In other words, to motivate the agent, the principal needs to adopt a more lenient project implementation policy, either by urging the manager to be less stringent in his recommendation or by direct intervention.¹⁰

We denote the former type of contracts as *mediation* contracts—the manager mediates between the principal and the agent by sometimes pooling a “bad” project with the rest. In contrast, we denote the latter type of contract as *intervention* contracts—the principal, with some probability, may intervene and implement the project by eschewing the manager's evaluation.

From [Lemma 1](#) we know that whenever a mediation or an intervention contract is sustained in equilibrium, it is optimal. The next lemma shows that if neither of them can be sustained in equilibrium, then there is no other equilibrium with production. (In the Appendix, the proof of this lemma is presented after the proofs of [Proposition 2](#) and [Proposition 3](#) as it uses some conditions that are discussed later in the text).

Lemma 2. *Suppose $(\alpha, \sigma) \in B$. If neither the mediation contract nor the intervention contract can be sustained as an equilibrium, then the only equilibrium is one in which all players take the outside option in all periods.*

In what follows, we limit attention to these two types of contracts and analyze the conditions under which they are sustained in equilibrium. Notice that the two contractual forms represent two distinct ways of managing the middle manager's excessive loyalty (to the top). Under a mediation contract, the manager is induced to act as if he were more aligned with the interests of the frontline worker than he actually is. By contrast, under an intervention contract, some of the times, the principal directly intervenes to counterbalance the manager's excessive loyalty.

5 Mediation and intervention contracts

5.1 Mediation contract

Under a *mediation* contract, the manager reports $r = 1$ when he sees $(\omega, s) \neq (C, 0)$. And when $(\omega, s) = (C, 0)$, he reports $r = 1$ if and only if the outcome of the randomizing device $z < \eta^*$, i.e., he reports $r = 1$ with probability η^* . Furthermore, the principal always follows the recommendation of the manager.

¹⁰The role of strategic leniency in incentive provision has been recognized in other contexts as well; see, e.g., Levitt and Snyder (1997) and Fairburn and Malcomson (2001).

Recall that the manager would prefer to have a project canceled if he sees $(\omega, s) = (C, 0)$, and can indeed ensure cancellation of the project by reporting $r = 0$. But if the manager deviates and reports $r = 0$ when $z < \eta^*$, the agent triggers punishment by terminating the relationship. Thus, for such a contract to be feasible, the manager's on-path payoff in the game must be at least as large as his deviation payoff, which is his outside option. The following incentive constraint for the manager stipulates this condition:

$$(1 - \delta) \Pr(V_M = y \mid \omega = C, s = 0) y + \delta v^* \geq v_0,$$

i.e.,

$$(1 - \delta) \left(\frac{1 - \sigma}{2 - \sigma} \right) y + \delta v^* \geq v_0. \quad (IC_M^M)$$

On the other hand, the principal must have an incentive to implement the project when $r = 1$ even though such a message pools some unprofitable projects for her (when $s = 0$) with the profitable ones (when $s = 1$). Should she deviate, the manager triggers punishment by terminating the relationship, and the principal obtains her outside option. Therefore, the following incentive constraint must hold:

$$(1 - \delta) \Pr(V_P = Y \mid r = 1) Y + \delta \pi^* \geq \pi_0,$$

i.e.,

$$(1 - \delta) \frac{\frac{1}{2}\sigma + \frac{1}{2}(1 - \sigma)(1 - \alpha(1 - \eta^*))}{1 - \alpha(1 - \eta^*)(1 - \frac{1}{2}\sigma)} Y + \delta \pi^* \geq \pi_0. \quad (IC_P^M)$$

We can now characterize the conditions under which the mediation contract may be sustained as a PBE of the game. Note that from the definition of η^* , it follows that

$$\alpha_1(\sigma) = \alpha(1 - \eta^*) = \frac{1 - c/b}{1 - \sigma/2}, \quad (1)$$

where $\alpha_1(\sigma)$ is as defined in **Proposition 1**: for any $\sigma \in [\sigma_0, 1]$, $\alpha_1(\sigma)$ is the maximal loyalty level of the manager for which the agent still can be incentivized to exert effort in the static game (i.e., for a given σ , (IC_A) binds at $\alpha_1(\sigma)$).

Proposition 2. (Mediation contract) *There exist $\delta_M \in (0, 1)$ and $\alpha_M(\sigma, \delta) \geq \alpha_1(\sigma)$ such that for any $\delta \geq \delta_M$, the payoffs (v^*, π^*) can be sustained by the mediation contract if $(\alpha, \sigma) \in B_M := \{(\alpha, \sigma) \mid \alpha_1(\sigma) \leq \alpha \leq \alpha_M(\sigma, \delta), \sigma \geq \sigma_0\}$. In particular, the following holds for the mediation contract:*

- (i) *When $(\omega, s) = (C, 0)$, the manager reports $r = 1$ with probability η^* and reports $r = 0$ with probability $1 - \eta^*$; when $(\omega, s) \neq (C, 0)$, the manager reports $r = 1$. The principal implements the project if and only if $r = 1$.*
- (ii) *The threshold $\alpha_M(\sigma, \delta)$ increases in δ but may be increasing or decreasing in σ , depending on the value of δ :*
 - (a) *If $\delta_M < \hat{\delta} := b/(2b - c)$, then $\alpha_M(\sigma, \delta)$ decreases in σ if $\delta_M \leq \delta \leq \hat{\delta}$ and increases in σ if $\delta \geq \hat{\delta}$.*

(b) If $\delta_M > \hat{\delta}$, $\alpha_M(\sigma, \delta)$ always increases in σ for all $\delta \geq \delta_M$.

Proposition 2 (illustrated in Figure 3 below) shows that when players are sufficiently patient, the mediation contract may indeed be the optimal relational contract provided that the manager is not too loyal to the principal. The threshold on the loyalty, α_M , is increasing in the discount rate δ . The impact of the manager's skill level (σ) on the loyalty threshold is more nuanced: the threshold decreases in σ if δ is low but increases in σ otherwise.

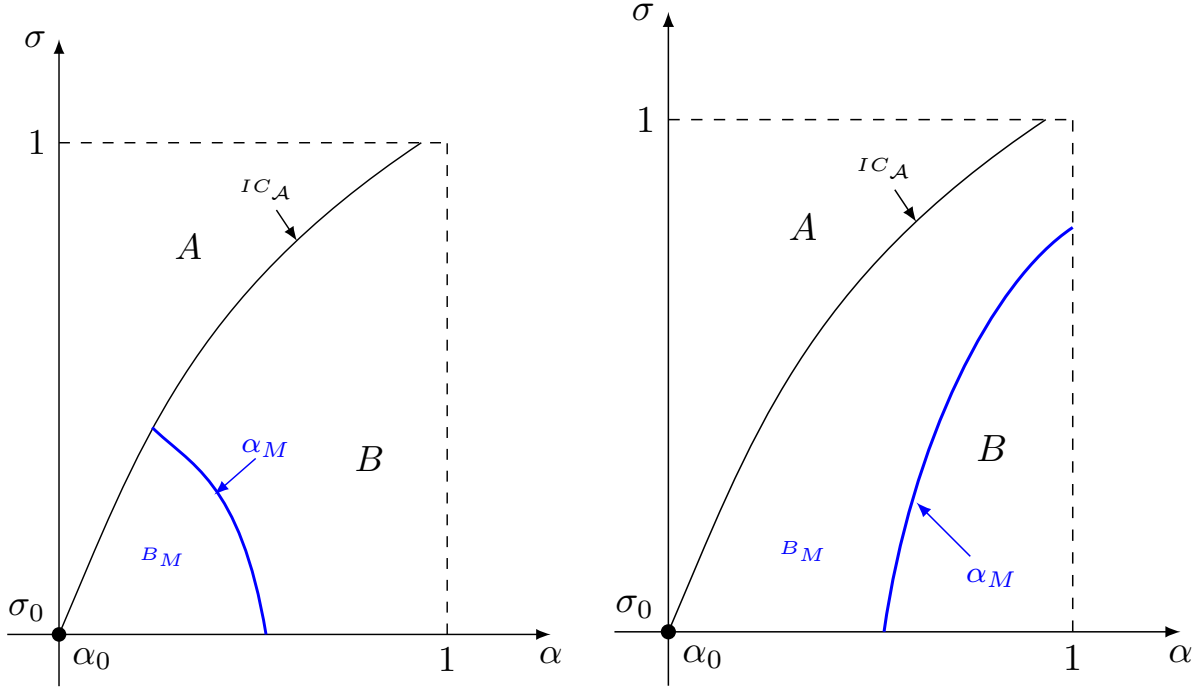


Figure 3: Mediation can be sustained as the optimal relational contract in region B_M .
Left panel: $\delta_M < \hat{\delta}$ and $\delta \in (\delta_M, \hat{\delta})$; right panel: $\delta > \max\{\hat{\delta}, \delta_M\}$.

The intuition for this result is as follows. Recall that we are focusing on a scenario where $(\alpha, \sigma) \in B$; i.e. in the static game, the agent's incentive constraint (IC_A) is violated and no project gets generated. The agent correctly anticipates that the manager, given his skill and loyalty, would be too discerning in his recommendation for project implementation, and the agent's expected benefit from the project would not cover his cost of effort. To encourage the agent to exert effort, a mediation contract props up the likelihood of project implementation as the manager recommends some projects that he would otherwise prefer to cancel. The manager is expected to behave as if his loyalty to the principal, α , were lowered to $\alpha_1(\sigma)$. In other words, the relational incentives drive a wedge between the manager's real loyalty (α) and his "revealed" loyalty ($\alpha_1(\sigma)$) to the principal when it comes to reporting on the project.

An immediate implication of this observation is that the principal's incentive constraint, (IC_P^M), is always satisfied: under such a contract, the principal's payoff in each period is the same as her payoff in the static model if the manager's loyalty were $\alpha_1(\sigma)$, and, by (IC_P), for any $\sigma \geq \sigma_0$ this

payoff is at least π_0 (i.e., the point $(\alpha_1(\sigma), \sigma)$ lies on (IC_A) where (IC_P) is always satisfied). Since (IC_P^M) is always satisfied, a *mediation* contract can be supported in a PBE as long as it satisfies the manager's incentive constraint (IC_M^M) .

Consider the manager's on-path payoff as given by the left-hand side of (IC_M^M) . One may readily note that for any $(\alpha, \sigma) \in B$ that is sufficiently close to (IC_A) , the associated value of η^* is also small. Hence, the manager's continuation payoff, v^* , is sufficiently close to his static game payoff, which is always larger than his outside option v_0 . Thus, for any such $(\alpha, \sigma) \in B$, the manager's continuation payoff $v^* > v_0$, and therefore, (IC_M^M) can be met if δ is sufficiently large. Conversely, for any δ large enough, there exists a set of (α, σ) that satisfies this constraint. To characterize this set, we need to explore how these parameters affect the manager's current and continuation payoffs.

The manager's current period payoff (when he recommends a bad project) decreases in σ but is independent of α . The signal $s = 0$ is a stronger indication of low value ($V_M = 0$) when the signal is more precise (i.e., when σ is high). But as the manager's current period payoff reflects his expected value of the project conditional on the state being C , it is not affected by the prior distribution (α) over the states.

The manager's continuation payoff, v^* , however, is affected by both σ and α : it is increasing in σ and decreasing in α . A higher precision of the signal means that ex-ante, the manager is more likely to get his preferred projects implemented and it is less likely that a project gets implemented when he prefers to have it canceled. That is, a higher σ moves probability from the state-signal pair $(\omega, s) = (C, 0)$ where the manager gets 0 from the project to the other three pairs where the manager gets a benefit of y . But a higher loyalty, α , moves probability from state D to state C . In particular, the ex-ante likelihood of obtaining the state-signal pair $(\omega, s) = (C, 0)$ increases and that of $(\omega, s) = (D, 0)$ decreases. That is, it becomes more likely that the manager would find himself in a scenario where he prefers to cancel the project but it still gets implemented, and it becomes less likely that a project that may be deemed valuable to the manager would be implemented.¹¹

As α always reduces the manager's on-path payoff, we readily see that for any given σ and (large enough) δ , there is a cutoff value of α , α_M , for which (IC_M^M) binds and is violated if and only if $\alpha > \alpha_M$. To see how α_M depends on σ , note that since σ has opposing effects on the manager's current and continuation payoffs, its net effect would depend on δ , the relative weight on the two payoffs. When δ is relatively small, σ 's negative effect on the current payoff dominates its positive effect on the continuation payoff, and the manager's on-path payoff decreases in σ . Consequently, for (IC_M^M) to bind, α_M must fall. But the opposite holds when δ is sufficiently large. As the manager's on-path payoff now increases in σ , for (IC_M^M) to bind, α_M must increase as well.

Thus, in a mediation contract, the manager is asked to behave as if he were less loyal to the principal. But the more loyal the manager truly is, the more costly it is for him to do so. Consequently, there is an upper bound on the manager's loyalty beyond which mediation cannot be sustained.

¹¹Note that when $s = 1$ the project is always valuable and is surely implemented, but the probability of obtaining this signal does not depend on α .

5.2 Intervention contract

We now explore the feasibility of an *intervention* contract. Under the intervention contract, the principal may sometimes decide on implementation by completely ignoring the manager's report. In particular, if the outcome of the public randomizing device $x < \eta^*$, she always implements the project disregarding the manager's recommendation, and anticipating that, the manager sends a completely uninformative report. If instead $x \geq \eta^*$, the principal and the manager behave as in the static game: the manager reports $r = 0$ when $(\omega, s) = (C, 0)$ and reports $r = 1$ otherwise, and the principal implements the project if and only if $r = 1$.

Notice that in an intervention contract, the manager does not have any profitable deviation from his reporting strategy. In the case where the principal is expected to ignore the manager's report, he cannot do any better than babbling. And in the case where the principal adheres to the manager's report, the manager's equilibrium reporting strategy calls for the implementation of the project only when it is indeed profitable for the manager to do so. Thus, we only have to ensure that the manager has the incentive to participate in this contract, i.e.,

$$v^* \geq v_0. \quad (IC_M^I)$$

The principal, on the other hand, may be tempted to cancel a project when he is expected to proceed with it. There are two cases to consider. First, if $x < \eta^*$ the principal must implement any project that the agent generates even though the project's (prior) expected value is less than the principal's outside option. Thus, the principal's on-path payoff must be larger than his deviation payoff:

$$(1 - \delta) \Pr(V_P = Y)Y + \delta\pi^* = (1 - \delta)\frac{1}{2}Y + \delta\pi^* \geq \pi_0. \quad (IC_P^I)$$

Second, when $x \geq \eta^*$, the principal must have incentives to adhere to the manager's recommendation who calls for implementation except when $(\omega, s) = (C, 0)$. Thus, we have the following constraint:

$$(1 - \delta) \Pr(V_P = Y \mid (\omega, s) \neq (C, 0))Y + \delta\pi^* \geq \pi_0. \quad (IC_P^{I'})$$

One may, however, readily note that (IC_P^I) implies $(IC_P^{I'})$, and hence, $(IC_P^{I'})$ can be ignored in our subsequent analysis. The proposition below characterizes the conditions under which an intervention contract may be sustained as a PBE of the game.

Proposition 3. (Intervention contract) *There exist $\delta_I \in (0, 1)$, cutoffs $\alpha_I(\sigma) \in [\alpha_1(\sigma), 1]$, and $\sigma_I(\delta) \in [\sigma_0, 1]$ such that for any $\delta \geq \delta_I$, the payoff (v^*, π^*) can be sustained by the intervention contract if $(\alpha, \sigma) \in B_I := \{(\alpha, \sigma) \mid \alpha_1(\sigma) \leq \alpha \leq \alpha_I(\sigma), \sigma_I(\delta) \leq \sigma \leq 1\}$. In particular, the following holds for the intervention contract:*

- (i) *If $x \leq \eta^*$, the principal implements any project that the agent generates and the manager sends a report that is completely uninformative. Otherwise, the manager reports $r = 0$ when $(\omega, s) = (C, 0)$ and reports $r = 1$ otherwise. The principal implements the project if and only if $r = 1$.*

- (ii) The cutoff $\alpha_I(\sigma)$ increases in σ and the cutoff $\sigma_I(\delta)$ decreases in δ ; furthermore, $\sigma_I(\delta_I) = 1$ and $\sigma_I(\delta) = \sigma_0$ as $\delta \rightarrow 1$.

Proposition 3, illustrated in **Figure 4** below, closely parallels its counterpart on mediation (**Proposition 2**). When the players are sufficiently patient, intervention is an optimal relational contract provided that the manager is not too loyal to the principal but is sufficiently skilled. The skill threshold decreases in δ , whereas the loyalty threshold α_I is independent of δ but increases in σ .

The intuition for this result also bears close resemblance to that of **Proposition 2**, though there are a few subtle differences. First, the principal's incentive constraint (IC_P^I) may bind as the princi-

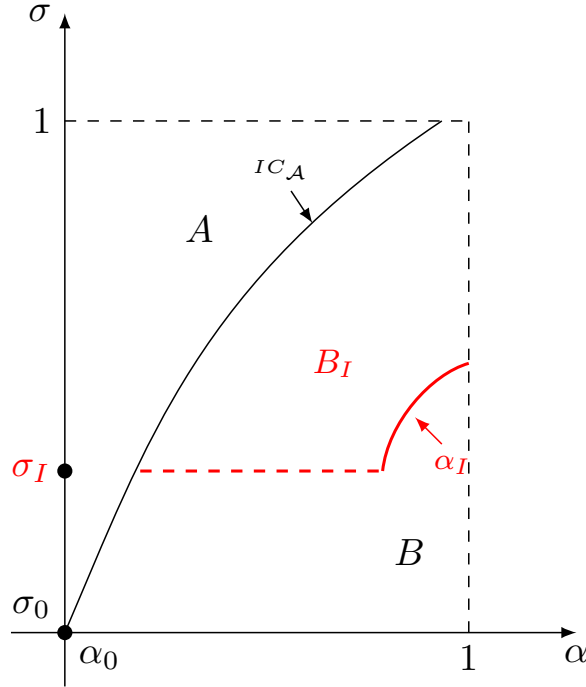


Figure 4: Intervention can be sustained as the optimal relation contract in region B_I .

pal would prefer to reject any project that the agent may generate when she lacks further information about its quality. As she implements the project ignoring the manager's report, the precision of the manager's signal, σ , does not affect her current period payoff whereas her continuation payoff is increasing in σ . Thus, the left-hand side of (IC_P^I) increases with σ . Furthermore, the principal's continuation payoff π^* is independent of α as the state affects the principal's payoff only through the ex-ante probability of project implementation, which remains constant when (IC_A^*) binds (i.e., η^* , the probability of implementation through principal's intervention, adjusts with α to ensure $\alpha(1 - \eta^*)$ stays fixed at $\alpha_1(\sigma)$). Thus, the principal's on-path payoff, the left-hand side of (IC_P^I) is increasing in σ but unaffected by α .

This observation, combined with the fact that the on-path payoff is also increasing in δ , implies that for any δ not too low, (IC_P^I) would be satisfied if σ is above a threshold $\sigma_I(\delta)$, and this

threshold decreases with δ . Thus, the intervention contract remains feasible as long as it ensures the manager's participation, i.e., it satisfies (IC_M^I) . And as the manager's continuation value v^* is decreasing in α , there is a threshold α_I (given σ) such that (IC_M^I) is satisfied if and only if $\alpha \leq \alpha_I(\sigma)$.

That $\alpha_I(\sigma)$ increases in σ can be gleaned from the fact that $\alpha_I(\sigma)$ is the value of α (given σ) for which (IC_M^I) binds (see Lemma 1), and v^* is increasing in σ and decreasing in α . As v^* is independent of δ , so is α_I .

Thus, in an intervention contract, the principal sometimes implements the project without consulting the manager, incurring an expected loss. This can be optimal only if she obtains sufficiently high value when she does consult him. Consequently, there is a lower bound on the manager's skill below which intervention becomes unsustainable.

6 Optimal relational contract

In light of the mediation and intervention contracts discussed above, we can now explore how the manager's characteristics affect the optimal relational contract.

Theorem 1. (Optimal relational contract) Fix Y/π_0 , and let $\delta \geq \delta_I$ so that $B_I \neq \emptyset$. Then, there exist two cutoffs τ_1 and τ_2 ($> \tau_1$) such that:

- (i) If $y/v_0 < \tau_1$, then $B_M = \emptyset$, i.e., the mediation contract cannot be sustained.
- (ii) If $\tau_1 \leq y/v_0 < \tau_2$, then $B_M \setminus B_I \neq \emptyset$ and $B_I \setminus B_M \neq \emptyset$, i.e., there exist parameter regions where the optimal contract is sustained only by the mediation contract and parameter regions where the optimal contract is sustained only by the intervention contract.
- (iii) If $y/v_0 \geq \tau_2$, then $B_I \subseteq B_M$, i.e., if the optimal relational contract can be sustained by the intervention contract, it can also be sustained by the mediation contract.

The above theorem indicates that the optimality of the mediation contract critically hinges on the ratio of the manager's project value to his outside option (y/v_0). Only the intervention contract may be sustained when this ratio is sufficiently small (part (i)). In contrast, when this ratio is sufficiently large, mediation dominates intervention—if intervention is optimal then so is mediation, but mediation could be sustained as a PBE even when the intervention contract remains infeasible (part (iii)). However, the most interesting case arises when this ratio is at an intermediate range. Here, depending on the parameter values, either the intervention or the mediation contract may emerge as the *unique* optimal relational contract (part (ii); Figure 5 illustrates this case).

To see the intuition for this result, consider part (ii) first. Recall that the manager's continuation value of the relationship (v^*) is increasing in his skill but decreasing in loyalty, whereas the principal's continuation value (π^*) increases in the manager's skill (σ) but does not depend on his loyalty (α). When the manager is not very skilled, the principal's continuation value of the relationship is low, but the manager's continuation value may still be large provided he is not quite

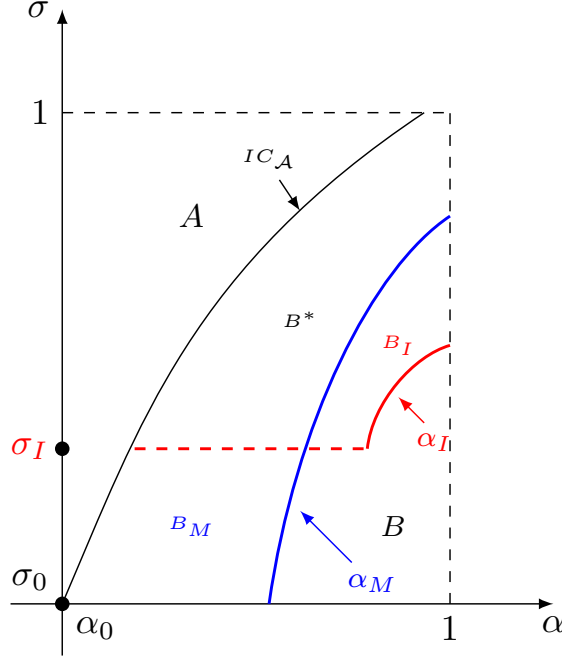


Figure 5: Optimal contract for high δ (i.e., $\delta > \max\{\delta_M, \delta_I\}$) in case (ii):
In region B_M only mediation is feasible; in B_I only intervention is feasible; in region B^* both modes are feasible.

loyal to the principal. Thus, the threat of termination by the agent may deter the manager from deviating but would not deter the principal. Consequently, only the mediation contract may be sustained. But for a highly skilled manager with high loyalty to the principal, the opposite holds. His continuation value may be too low to deter deviation whereas the principal may have a high continuation value and would not find it profitable to deviate.¹² Thus, when the manager's skill (σ) and loyalty (α) are not too high, mediation may be the unique optimum, and when the manager is highly skilled and relatively loyal, intervention may be the unique optimum.

This result shows how loyalty within the organization can be optimally managed. If the manager's skill and loyalty are both relatively low, mediation is optimal. In equilibrium, while recommending implementation, the manager sometimes pools a "bad" project (one that is likely to yield low value to both the manager and the principal) with "good" ones (one that is sure to yield high value to the manager), and the principal always accepts the manager's recommendation. In contrast, if the manager's skill and loyalty are relatively high, an intervention contract becomes optimal. In such a contract, some of the time, the principal intervenes and directly implements the project, ignoring the manager's report. And on other times, the manager recommends only those projects that are profitable to him, and the principal abides by his recommendation.

The above result also highlights how the mode of communication and decision-making in an

¹²It is also worthwhile to note that $B_M = B_I = \{(\alpha, \sigma) \mid \alpha_1(\sigma) \leq \alpha \leq \alpha_I(\sigma), \sigma_0 \leq \sigma \leq 1\}$ as $\delta \rightarrow 1$. As the players put little weight on their deviation gains in the current period, both mediation and intervention contracts can be sustained as long as the players' continuation payoffs exceed their respective outside options, i.e., $(\pi^*, v^*) \geq (\pi_0, v_0)$.

organization may be driven by the characteristics of its middle managers. Following the seminal work by Bertrand and Schoar (2003), a large literature has documented a strong influence of manager fixed effects in corporate practices or “management styles”—i.e., the pattern of decision-making—observed across organizations. The variations in styles are typically attributed to the variations in the top management’s preferences, skills, or biases in their opinions (Rotemberg and Saloner, 1993; Bertrand and Schoar, 2003; Dessein and Santos, 2021). An important implication of this finding is that the management style that the headquarters adopts could stem from the personal characteristics of the middle managers rather than that of the CEO herself.

To see the intuition for parts (i) and (iii), it is useful to make the following observation in relation to the feasibility of the mediation contract (as given in [Proposition 2](#)).

Lemma 3. δ_M is decreasing in y/v_0 and $\alpha_M(\sigma, \delta)$ is increasing in y/v_0 .

The larger the project value to outside option ratio for the manager, the more he is likely to gain from continuing the relationship, and the smaller is his temptation to renege on the mediation contract. Thus, δ_M , the minimum value of δ that sustains the mediation contract, is decreasing in y/v_0 . This observation readily implies the result in part (i): given δ , if y/v_0 is sufficiently small, the mediation contract is not sustainable. Finally, recall that the mediation contract is sustainable if and only if $\alpha \leq \alpha_M(\sigma, \delta)$. As noted above, when the middle manager gains more from continuing the relationship, he is more willing to follow the mediation contract. Hence, as stated in part (iii), an increase in y/v_0 increases the threshold $\alpha_M(\sigma, \delta)$, expanding the range where the mediation contract is optimal.

The above result highlights the role of the manager’s loyalty *to the organization* as captured by his project value to outside option ratio in the mediation contract. In particular, the manager can successfully mediate between the bottom and the top rungs of the organization only if he is sufficiently loyal to the organization. Furthermore, in the optimal mediation contract, his loyalty to the organization complements his loyalty to the top: a manager who is keenly loyal to the top can still succeed in his dual role (of incentivizing the workers at the bottom while maintaining credibility at the top) if he is also sufficiently loyal to the organization. In other words, the optimal management of loyalty *within* the organization may depend on the manager’s loyalty *to* the organization.

We conclude this section with the following two remarks. First, in our model of hierarchical organization, a relational contract consists of three bilateral but interconnected relations among the agent, manager, and principal. When the mediation contract is optimal, it is the relational contract between the manager and the agent that ensures the provision of effort incentives, whereas, under the intervention contract, the relational contract between the principal and the agent comes into play. Thus, our finding highlights that depending on the characteristics of the manager, the provision of incentives may rely on different bilateral relations within the organization.

Second, recall that for the mediation contract to be feasible, the principal may not observe the realization of the randomizing device z , as otherwise, it need not be sequentially rational for the principal to always follow the manager’s recommendation. Thus, our result indicates that it may be beneficial for the principal to keep the relationship between the agent and the manager at arm’s

length. The principal may undermine incentives in the organization by micromanaging the relational contract between the agent and the manager where the principal demands to see the realization of z . In particular, we have the following corollary.

Corollary 1. *Suppose that the randomizing device z is public. Then, for any $(\alpha, \sigma) \in B \setminus B_I$, the only equilibrium is one in which all players take the outside option in all periods.*

That is, when z is public, if the optimal relational contract can be sustained by the mediation contract, it can also be sustained by the intervention contract.

7 Discussion and conclusion

7.1 Discouraging implementation

Our analysis of the baseline model shows that the middle manager may also lose his effectiveness when he is too close to the agent as he recommends implementation too frequently. This type of distortion arises in regions A and D in [Figure 1](#). To ameliorate this problem in a long-term relationship, the parties must discipline the manager to recommend projects in a way that is more aligned with the principal's interests.

The key, then, is to incentivize the manager to forgo some of the projects that bring him private benefits but are unprofitable for the principal (i.e., projects with $\omega = D$ and $s = 0$).¹³ Since the principal's realized value from the project (when implemented), V_P , is publicly observed, the principal can discipline the manager by threatening to terminate the relationship if the manager's recommended project results in a poor outcome. However, the principal may need to exercise some leniency to prevent the manager from becoming too cautious in making recommendations, which could blunt the agent's incentives to exert effort. To strike this balance, the punishment for a poor outcome may be conditioned on the realization of the public randomization device, x .

In what follows, we call an action profile of the stage game " β -lenient" if the agent exerts effort, the principal follows the manager's recommendation, and the manager reports $r = 1$ if $s = 1$, $r = 0$ if $(\omega, s) = (C, 0)$, and when $(\omega, s) = (D, 0)$ he reports $r = 1$ if $x \leq \beta$ and $r = 0$ if $x > \beta$. The principal terminates the relationship only if $V_P = 0$ and $x > \beta$. That is, with probability β , the manager may recommend a "bad" project (i.e., $s = 0$) as he is assured of leniency should the project yield a poor outcome for the principal. Note that, when $\beta = 1$, the pattern of project implementation coincides with that in the baseline model ([Proposition 1](#)); and when $\beta = 0$, the action profile indicates that only the "good" project (i.e., $s = 1$) can get implemented.

Proposition 4. *Suppose $(\alpha, \sigma) \in \{(\alpha, \sigma) \mid \sigma \geq \sigma_0 \text{ and } \alpha < \alpha_1(\sigma)\}$.*

¹³In region D , the manager can mediate between the principal and the agent by using the randomizing device z_t , as is the case in the mediation contract discussed in [Section 5.1](#). Since the manager can pool the "bad" projects (those with $(\omega, s) = (D, 0)$) with the "good" projects (those with $s = 1$) without the principal being aware, the principal's incentive constraint will be trivially satisfied. Hence, we can focus on the manager's incentives.

1. There exist $\bar{\delta}$ (which depends on α and σ) such that if $\delta \in [\bar{\delta}, 1)$, the stationary contract, where the parties' action profile is β -lenient in every period (for some $\beta \in [0, 1]$), is optimal.
2. If $\sigma > \sigma_0$, then there exists $\underline{\delta} \in (0, \bar{\delta})$ (which depends on α and σ) such that if $\delta \in [\underline{\delta}, \bar{\delta})$, the optimal relational contract is non-stationary and can be the following:
 - (a) In the first period, the parties start the relationship by playing a β -lenient action for some $\beta \in [0, 1]$.
 - (b) In the subsequent periods, their action randomizes between β -leniency and β' -leniency for some $\beta' \in (\beta, 1]$.
 - (c) Once the β' -lenient action is realized, the future actions are always β' -lenient.

(The proof of the above proposition relies on techniques that closely resemble the methods developed for the characterization of the PPE payoff as in Abreu et al. (1990), and is given in the Online Appendix.) **Proposition 4** shows that the optimal relational contract allows for leniency while using termination for poor project outcome, and the contract dynamics depend on the players' discount factor. Recall that in Regions *A* and *D*, any improvement over the static equilibrium requires the manager to sometimes ignore his short-term private benefit and recommend implementation in line with the principal's interest. In the dynamic relationship, the principal fulfills the requirement by promising the manager a high enough continuation payoff if he obeys and by threatening him with termination if he deviates. Since the principal's optimal action (among the individually rational actions) can provide the manager with a payoff strictly higher than v_0 , such an action can be sustained as an equilibrium as long as δ is sufficiently close to one.

However, when δ is not sufficiently high, sustaining the principal's optimal action on the equilibrium path counts on back-loading the incentives for the manager. Specifically, the principal must promise the manager higher future leniency to enforce more stringent actions in the earlier periods, while maintaining the agent's incentive to exert effort. This can be fulfilled as long as the manager's skill is above σ_0 . In the long run, the relationship falls into the action with relatively higher leniency. Such back-loaded incentives and long-run dynamics are reminiscent of the optimal contracts explored in the recent literature on relational incentives with private information (see, e.g., Ray, 2002, Chassang, 2010, and Li et al., 2017).

7.2 An alternative information structure for the manager's signal

In our model, the middle manager always identifies a bad project for the principal with certainty but detects a good project only some of the time. Thus, the "skill" of the manager refers to his ability to identify a good project as such. Such a "good news is definitive" feature of the signal process helps in improving the analytical tractability of the model, but it is not essential for delivering its key insights on the value of "moderate" loyalty and how the range of moderation expands with the manager's skill (**Proposition 1**).

To see this, consider the following alteration to our baseline model: as before, assume that once a project is generated the manager observes its type (ω) and the signal s on its value to the principal, but the informativeness of the manager's signal is given now as:

$$\Pr(s = 1 \mid V_{\mathcal{P}} = Y) = \Pr(s = 0 \mid V_{\mathcal{P}} = 0) = \sigma > \frac{1}{2}.$$

That is, neither of the two signals is definitive but $s = 1$ is better news about the project than $s = 0$ with σ being the precision of the signal. We continue to focus on a parameter region where informative communication can be supported in equilibrium, and, analogous to Assumption 1, we impose the following parametric restrictions:

Assumption 1'. (i) $(1 - \sigma)y < v_0 < \sigma y$ and (ii) $(1 - \sigma)Y < \pi_0 < \sigma Y$.

As in our baseline model, in any equilibrium, the manager will always recommend the project when $\omega = D$ or $(\omega, s) = (C, 1)$, as he obtains an expected project value of y or σy , respectively, both exceeding his outside option v_0 (by Assumption 1'). Thus, in any equilibrium where the manager's report is informative, it must be that $r = 0$ if and only if $(\omega, s) = (C, 0)$. And for such a reporting strategy to be sequentially rational for the manager, it must be that

$$v_0 > \mathbb{E}[V_{\mathcal{M}} \mid (\omega, s) = (C, 0)] = (1 - \sigma)y.$$

which also holds under Assumption 1'.

Next, consider the incentive constraints of the principal that ensure that she follows the manager's recommendation: $\Pr(V_{\mathcal{P}} = Y \mid r = 0)Y < \pi_0 \leq \Pr(V_{\mathcal{P}} = Y \mid r = 1)Y$, i.e.,

$$(1 - \sigma)Y < \pi_0 \leq \frac{1 - \alpha(1 - \sigma)}{2 - \alpha}Y, \quad (IC_{\mathcal{P}}^*)$$

and the constraint on the agent that ensures he exerts effort to generate a project, i.e.,

$$\Pr(r = 1 \mid e = 1)b = \left(1 - \frac{1}{2}\alpha\right)b \geq c. \quad (IC_{\mathcal{A}}^*)$$

Thus, a PBE with informative communication exists if and only if $(IC_{\mathcal{P}}^*)$ and $(IC_{\mathcal{A}}^*)$ hold, and these two constraints stipulate a lower and upper bound on α (observe that under Assumption 1', only the right-hand side inequality in $(IC_{\mathcal{P}}^*)$ may be binding). Hence, we again obtain the result that for a middle manager to be effective, his loyalty to the top must lie in a moderate range. Moreover, as noted in our baseline model, this range is increasing in the manager's skill as $(IC_{\mathcal{P}}^*)$ is relaxed when σ increases.

7.3 Conclusion

Middle managers play a pivotal role in linking the top management with the frontline workers, and their alignment of loyalty within the organizational hierarchy may affect the organizational

performance. This issue is particularly salient in innovative organizations where the middle managers undertake a critical dual role: as conduits of information between headquarters and employees “on the ground,” they must credibly convey decision-relevant information upward while simultaneously keep junior employees motivated by promoting their ideas to firm leadership and securing corporate support for their work. Thus, a middle manager must be skilled in distinguishing promising ideas from unpromising ones and in navigating the delicate balance of dual loyalty—maintaining credibility with both the top and bottom levels of the organization.

Our analysis illustrates that in hierarchical organizations, a balanced alignment of loyalty may be critical for overall firm performance. Middle managers are most effective when their loyalty to top falls within a moderate range—strong enough to maintain credibility upward, yet limited enough to preserve trust below. This range expands with the manager’s skill; in other words, more competent managers can sustain credibility across a wider range of loyalty levels.

Another important insight from our analysis is that loyalty within the organization can be managed through the carefully designed communication and decision-making protocols. In a long-term employment relationship, we show how relational incentives can be used to manage loyalty and restore efficiency. When excessive loyalty to top management compromises organizational performance, two classes of relational contracts emerge as optimal. In *mediation* contracts, incentives induce the manager to behave as if he were less loyal, while in *intervention* contracts, the principal counterbalances excessive loyalty by occasionally disregarding the manager’s recommendations and making decisions that directly benefit the workers. However, mediation contracts are feasible (and optimal) only if the middle manager is sufficiently loyal to the organization.

Our findings have broader implications for understanding how loyalty both within and toward the organization affects organizational design and performance. First, the optimal management of internal loyalty depends on the interaction between these two dimensions of loyalty. Second, loyalty within and toward the organization jointly determine the firm’s patterns of communication and decision-making, thereby shaping the leadership’s managerial style.

Appendix

This Appendix contains the proofs omitted in the text.

Proof of Proposition 1. Step 1. Consider any equilibrium in which the principal’s project implementation policy depends on the manager’s report r , i.e., an equilibrium in which $\Pr(\iota = 1 \mid r = 1) \neq \Pr(\iota = 1 \mid r = 0)$. Suppose, without loss of generality, that $\Pr(\iota = 1 \mid r = 1) > \Pr(\iota = 1 \mid r = 0)$. Since

$$\mathbb{E}[V_M \mid (\omega, s)] = y > v_0 \text{ for all } (\omega, s) \neq (C, 0),$$

in such equilibrium the manager necessarily reports $r = 1$ whenever $(\omega, s) \neq (C, 0)$. Similarly, since

$$\mathbb{E}[V_M \mid (C, 0)] = ((1 - \sigma)/(2 - \sigma))y < v_0,$$

the manager necessarily reports $r = 0$ when $(\omega, s) = (C, 0)$.

Step 2. Given the above reporting strategy, it is optimal for the principal to follow the manager's recommendation—i.e., the project implementation policy $\Pr(\iota = 1 \mid r = 1) = \Pr(\iota = 0 \mid r = 0) = 1$ is optimal—whenever (IC_P) and (IC'_P) are satisfied. And, given the above reporting and project implementation strategies, it is optimal for the agent to choose $e = 1$ when (IC_A) is satisfied. Thus, when (IC_P) , (IC'_P) and (IC_A) are satisfied, there exists an equilibrium in which $e = 1$, the manager recommends the project's implementation if $(\omega, s) \neq (C, 0)$ and cancellation if $(\omega, s) = (C, 0)$, and the principal follows the manager's recommendation, provided the principal's ex-ante payoff from this equilibrium is at least as large as his outside option.

Step 3. As noted in the text, (IC'_P) is always satisfied under Assumption 1(i). Hence, such an equilibrium exists if and only if (IC_P) and (IC_A) are satisfied and the principal's payoff exceeds π_0 . Observe that (IC_P) is equivalent to

$$\alpha \geq \frac{2\pi_0 - Y}{\sigma(Y - \pi_0) + (2\pi_0 - Y)} := \alpha_0,$$

and (IC_A) is equivalent to

$$\alpha \leq \frac{1 - \frac{c}{b}}{1 - \frac{1}{2}\sigma}.$$

Let $\alpha_1 := \min\{(1 - \frac{c}{b}) / (1 - \frac{1}{2}\sigma), 1\}$. It is routine to check that $\alpha_1 \geq \alpha_0$ if and only if

$$\sigma \geq \sigma_0 := \frac{2c(2\pi_0 - Y)}{2c(\pi_0 - Y) + Yb},$$

and that $\alpha_1 = \alpha_0$ if $\sigma = \sigma_0$. Also note that $\sigma_0 > 1$ if and only if $Y/\pi_0 < 2c/b$.

Step 5. When $\sigma < \sigma_0$, either (IC_P) or (IC_A) is violated. If (IC_P) is violated, there is no equilibrium where the principal's implementation policy depends on the manager's report. In any equilibrium in which the project implementation is independent of the manager's report, the project is always rejected. So, $e = 0$ and all parties obtain their outside options. Similarly, when (IC_A) is not satisfied, in all equilibria, $e = 0$ and all parties obtain their outside options. In both cases, the principal's payoff is π_0 .

Step 5a. As $Y/\pi_0 < 2c/b \Leftrightarrow \sigma_0 > 1$, from step 5 above, it implies that for any given σ , in all equilibria, $e = 0$ and all parties obtain their outside options.

Step 6. Now, suppose $Y/\pi_0 > 2c/b$, and hence, $\sigma_0 < 1$. For any $\sigma \geq \sigma_0$ and $\alpha \in [\alpha_0, \alpha_1]$ the principal's ex-ante payoff under the aforementioned strategy profile (i.e., $e = 1$, the manager reports cancellation if and only if $(\omega, s) = (C, 0)$, and the principal always follows the manager's recommendation) is

$$\pi = \Pr(r = 1)\mathbb{E}[V_P \mid r = 1] + \Pr(r = 0)\pi_0 \geq \pi_0$$

(the inequality follows from the fact that the (IC_P) implies that $\mathbb{E}[V_P \mid r = 1] \geq \pi_0$). Thus, the above strategy profile constitutes an equilibrium of the game.

If $\sigma \geq \sigma_0$ but $\alpha \notin [\alpha_0, \alpha_1]$ either (IC_P) or (IC_A) is violated, and as argued in Step 4 above, an equilibrium where $e = 1$ does not exist, and the parties take their outside option.

Finally, it is routine to check from the formula for α_0 and α_1 given in Step 3 that α_0 decreases in σ whereas α_1 increases in σ (and by Assumption 1 (iii), $\alpha_1(1) < 1$).

□

Proof of Lemma 1. Step 1. Proof of Part (i): We begin by showing that $\pi^* \geq \pi_0$. Decompose π^* as

$$\begin{aligned}
\pi^* &= \pi_0 + \Pr(r = 1)(\mathbb{E}(V_P \mid r = 1) - \pi_0) \\
&= \pi_0 + \Pr(r = 1)(\Pr(V_P = Y \mid r = 1)Y - \pi_0) \\
&= \pi_0 + \Pr(s = 1) \Pr(r = 1 \mid s = 1)(\Pr(V_P = Y \mid s = 1)Y - \pi_0) + \\
&\quad \Pr(s = 0) \Pr(r = 1 \mid s = 0)(\Pr(V_P = Y \mid s = 0)Y - \pi_0) \\
&= \pi_0 + \frac{1}{2}\sigma(Y - \pi_0) + \left(1 - \frac{1}{2}\sigma\right) (1 - \alpha(1 - \eta^*)) \left(\frac{1 - \sigma}{2 - \sigma}Y - \pi_0\right).
\end{aligned} \tag{2}$$

Now, from (1) we have $\alpha(1 - \eta^*) = \alpha_1$, and recall that $\alpha_1 \geq \alpha_0$ where α_1 and α_0 are as defined in Proposition 1. From these two observations and (2), it follows that

$$\begin{aligned}
\pi^* &\geq \pi_0 + \frac{1}{2}\sigma(Y - \pi_0) + \left(1 - \frac{1}{2}\sigma\right) (1 - \alpha_0) \left(\frac{1 - \sigma}{2 - \sigma}Y - \pi_0\right) \\
&= \pi_0,
\end{aligned}$$

where the equality is obtained by replacing α_0 with its expression as given in the proof of Proposition 1 (recall that by definition, at $\alpha = \alpha_0$, (IC_P) binds, i.e., $\Pr(V_P = Y \mid r = 1)Y - \pi_0 = 0$).

Step 2. Proof of Part (ii) and the first result in Part (iii):

Step 2a. We first claim that for any equilibrium, it is without loss to compute the manager's continuation payoff at any period t conditional on the agent's history $(h^{A,t})$ rather than the manager's $(h^{M,t})$. Formally, consider an arbitrary equilibrium, and fix an arbitrary period t . Recall that the private history of the agent up to date t is contained in that of the manager's. In what follows, with a slight abuse of notation, we denote this fact as $h^{A,t} \subset h^{M,t}$. Our claim can be represented as $v_t(h^{M,t}) = v_t(h^{A,t})$.

The claim follows from the following two observations. First, because the projects are i.i.d across periods, the manager's histories may affect his payoff only through their effect on his strategy. Second, note that the agent's strategy only depends on his private history $h^{A,t}$, the principal's strategy only depends on the public history $h^{P,t}$, and both $h^{A,t}$ and $h^{P,t}$ are part of the manager's history $h^{M,t}$ (as for any given period τ , $h^{P,\tau} \subset h^{A,\tau} \subset h^{M,\tau}$). Hence, given $h^{M,t}$, the manager's best response depends on $h^{A,t}$ only. In particular, such a response will give the manager a continuation payoff of $v_t(h^{A,t})$ for any history $\tilde{h}^{M,t}$ that satisfies $h^{A,t} \subset \tilde{h}^{M,t}$. Consequently, $v_t(h^{M,t}) = v_t(h^{A,t})$.

Step 2b. We now show that if $v^* < v_0$, the only equilibrium is one in which all players take their outside options; and if $v^* \geq v_0$, then v^* bounds the manager's continuation payoffs from above.

Consider an arbitrary equilibrium and the agent's continuation payoff $u_t(h^{A,t})$. In the equilibrium, for each $\tau > t$, denote $w_\tau := \Pr[e_\tau = 1 \mid h^{A,t}]$ the agent's period t -belief that he will exert effort at period τ , and as $p_\tau := \Pr[\iota_\tau = 1 \mid (\omega_\tau, s_\tau) \neq (C, 0), e_\tau = 1, h^{A,t}]$ and $q_\tau := \Pr[\iota_\tau = 1 \mid (\omega_\tau, s_\tau) = (C, 0), e_\tau = 1, h^{A,t}]$ the agent's belief about implementation conditional on $e_\tau = 1, \omega_\tau, s_\tau$, and his history $h^{A,t}$. Decompose $u_t(h^{A,t})$ as

$$\begin{aligned} u_t(h^{A,t}) &= (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau [\Pr((\omega_\tau, s_\tau) \neq (C, 0)) p_\tau b + \Pr((\omega_\tau, s_\tau) = (C, 0)) q_\tau b - c] \\ &= (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau \left[\left(1 - \left(1 - \frac{1}{2}\sigma\right)\alpha\right) p_\tau b + \left(1 - \frac{1}{2}\sigma\right)\alpha q_\tau b - c \right]. \end{aligned} \quad (3)$$

Similarly, decompose u^* as

$$u^* = (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau \left[\left(1 - \left(1 - \frac{1}{2}\sigma\right)\alpha\right) b + \left(1 - \frac{1}{2}\sigma\right)\alpha \eta^* b - c \right], \quad (4)$$

Since $b > 0, p_\tau \leq 1$ for all τ , and $u_t(h^{A,t}) \geq u^* = 0$ (recall that $u^* = 0$ follows from the definition of η^*), using (3) and (4) we obtain

$$\sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau q_\tau \geq \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau \eta^*. \quad (5)$$

Step 2c. Note that in this equilibrium,

$$\begin{aligned} v_t(h^{A,t}) &= v_0 + (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau \left[\Pr((\omega_\tau, s_\tau) \neq (C, 0)) p_\tau (y - v_0) + \right. \\ &\quad \left. \Pr((\omega_\tau, s_\tau) = (C, 0)) \alpha q_\tau (\mathbb{E}(V_{\mathcal{M}} \mid s = 0) y - v_0) \right] \\ &= v_0 + (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau \left[\left(1 - \left(1 - \frac{1}{2}\sigma\right)\alpha\right) p_\tau (y - v_0) + \right. \\ &\quad \left. \left(1 - \frac{1}{2}\sigma\right)\alpha q_\tau \left(\frac{1 - \sigma}{2 - \sigma} y - v_0\right) \right]. \end{aligned}$$

Using (5) and the fact that $y - v_0 > 0, p_\tau \leq 1$, and $\frac{1 - \sigma}{2 - \sigma} y - v_0 < 0$, it follows that

$$v_t(h^{A,t}) \leq v_0 + (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau k,$$

where

$$k := \left(1 - \left(1 - \frac{1}{2}\sigma\right)\alpha\right) (y - v_0) + \left(1 - \frac{1}{2}\sigma\right)\alpha \eta^* \left(\frac{1 - \sigma}{2 - \sigma} y - v_0\right).$$

Also note that v^* can be written as

$$v^* = v_0 + (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} k.$$

Now, $v^* < v_0 \Rightarrow k < 0$. Thus, if for any t the manager's continuation payoff $v_t(h^{\mathcal{M},t}) = v_t(h^{\mathcal{A},t}) \geq v_0$ we must have $w_\tau = 0$ for all $\tau > t$. That is, the agent never exerts effort in any future periods, and hence, the equilibrium is the one in which all players take their outside options (and $v_t = v_0$ for all t). On the other hand, $v^* \geq v_0 \Rightarrow k \geq 0$. In this case, as $w_\tau \leq 1$, $v^* \geq v_t(h^{\mathcal{A},t}) = v_t(h^{\mathcal{M},t})$.

Step 3. *Proof of second result in Part (iii):* Finally, we show that the principal's continuation payoff is bounded above by π^* for any equilibrium. Recall that this continuation payoff is written as $\pi_t(h^{\mathcal{P},t})$ for any period t .

Consider an arbitrary equilibrium. For each $\tau > t$, with a slight abuse of notation, let $p_\tau := \Pr[\iota_\tau = 1 \mid s_\tau = 1, e_\tau = 1, h^{\mathcal{A},t}]$ and $q_\tau := \Pr[\iota_\tau = 1 \mid s_\tau = 0, e_\tau = 1, h^{\mathcal{A},t}]$ the agent's belief of implementation conditional on $e_\tau = 1$ and s_τ . (As before, we continue to denote $w_\tau := \Pr[e_\tau = 1 \mid h^{\mathcal{A},t}]$ as the agent's period t -belief that he will exert effort at period τ).

Step 3a. We can write

$$u_t(h^{\mathcal{A},t}) = (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau \left[\frac{1}{2} \sigma p_\tau b + \left(1 - \frac{1}{2} \sigma \right) q_\tau b - c \right],$$

and

$$u^* = (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau \left[\frac{1}{2} \sigma b + \left(1 - \frac{1}{2} \sigma \right) (1 - \alpha(1 - \eta^*)) b - c \right].$$

Since $u_t \geq u^* = 0$, $b > 0$, and $p_\tau \leq 1$, we obtain

$$\sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau q_\tau \geq \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau (1 - \alpha(1 - \eta^*)). \quad (6)$$

Step 3b. Note that

$$\begin{aligned} \pi_t(h^{\mathcal{A},t}) &= \pi_0 + (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau \left[\frac{1}{2} \sigma p_\tau (Y - \pi_0) + \left(1 - \frac{1}{2} \sigma \right) q_\tau \left(\frac{1 - \sigma}{2 - \sigma} Y - \pi_0 \right) \right] \\ &\leq \pi_0 + (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} w_\tau \left[\frac{1}{2} \sigma (Y - \pi_0) + \right. \\ &\quad \left. \left(1 - \frac{1}{2} \sigma \right) (1 - \alpha(1 - \eta^*)) \left(\frac{1 - \sigma}{2 - \sigma} Y - \pi_0 \right) \right]. \end{aligned}$$

where the inequality follows from (6), and the facts that $\frac{1-\sigma}{2-\sigma} Y < \pi_0$, $Y - \pi_0 > 0$, and $p_\tau \leq 1$. Also

note that π^* can be written as

$$\pi^* = \pi_0 + (1 - \delta) \sum_{\tau \geq t+1} \delta^{\tau-t-1} \left[\frac{1}{2} \sigma (Y - \pi_0) + \left(1 - \frac{1}{2} \sigma \right) (1 - \alpha(1 - \eta^*)) \left(\frac{1 - \sigma}{2 - \sigma} Y - \pi_0 \right) \right].$$

Since $\pi^* \geq \pi_0$,

$$\frac{1}{2} \sigma (Y - \pi_0) + \left(1 - \frac{1}{2} \sigma \right) (1 - \alpha(1 - \eta^*)) \left(\frac{1 - \sigma}{2 - \sigma} Y - \pi_0 \right) > 0.$$

As $w_\tau \leq 1$, we obtain $\pi_t(h^{\mathcal{A},t}) \leq \pi^*$. Furthermore, because the principal's history is contained in that of the agent's, by the law of iterated expectation,

$$\mathbb{E} [\pi_t(h^{\mathcal{A},t}) \mid h^{\mathcal{P},t}] = \pi_t(h^{\mathcal{P},t}).$$

Thus, we obtain $\pi_t(h^{\mathcal{P},t}) \leq \pi^*$. □

Proof of Proposition 2. We begin by showing that the incentive compatibility constraint of the principal, $(IC_{\mathcal{P}}^M)$ is always satisfied. Recall that this constraint requires that

$$(1 - \delta) \Pr(V_{\mathcal{P}} = Y \mid r = 1)Y + \delta \pi^* \geq \pi_0.$$

Decomposing π^* based on the manager's report, we obtain

$$\pi^* = \Pr(r = 1) \Pr(V_{\mathcal{P}} = Y \mid r = 1)Y + \Pr(r = 0)\pi_0.$$

Since $\pi^* \geq \pi_0$ (by Lemma 1), this implies that $\Pr(V_{\mathcal{P}} = Y \mid r = 1)Y \geq \pi_0$, which in turn implies that $(IC_{\mathcal{P}}^M)$ is satisfied.

We now analyze the conditions under which the manager's incentive compatibility constraint $(IC_{\mathcal{M}}^M)$ is satisfied. From (1), we obtain that

$$\eta^* = 1 - \frac{1 - \frac{c}{b}}{\alpha \left(1 - \frac{1}{2} \sigma \right)}.$$

Furthermore, observe that we can write

$$\begin{aligned} v^* &= \left[\frac{1}{2} \sigma + \left(1 - \frac{1}{2} \sigma \right) (1 - \alpha) \right] y + \frac{1}{2} (1 - \sigma) \alpha \eta^* y + \left(1 - \frac{1}{2} \sigma \right) \alpha (1 - \eta^*) v_0 \\ &= \left(1 - \frac{c}{b} \right) v_0 + \left(\frac{1}{2 - \sigma} + \frac{c}{b} \left(\frac{1 - \sigma}{2 - \sigma} \right) - \frac{1}{2} \alpha \right) y, \end{aligned}$$

where the second equality follows from replacing η^* with its expression given above and simplifying.

In what follows, let $L_M(\alpha, \sigma, \delta)$ denote the left-hand side of (IC_M^M) . Differentiating L_M with respect to α , we obtain

$$\frac{\partial L_M}{\partial \alpha} = \delta \frac{\partial v^*}{\partial \alpha} = -\frac{1}{2}\delta y < 0.$$

Thus, L_M decreases with α . Now, note that when $(\alpha, \sigma) \in B$, $\alpha \geq \alpha_1(\sigma)$ and $\sigma \geq \sigma_0$. So, for any given $\sigma \geq \sigma_0$, the maximum value of L_M is obtained when $\alpha = \alpha_1(\sigma)$, and is given by

$$\begin{aligned} L_M(\alpha_1(\sigma), \sigma, \delta) &= (1 - \delta) \left(\frac{1 - \sigma}{2 - \sigma} \right) y + \delta v^*|_{\alpha = \alpha_1(\sigma)} \\ &= (1 - \delta) \left(\frac{1 - \sigma}{2 - \sigma} \right) y + \delta \left[\frac{c}{b}y + \left(1 - \frac{c}{b} \right) v_0 \right], \end{aligned}$$

where the second equality follows from using (1) and the expression of v^* above. Since $(1 - \sigma)/(2 - \sigma)$ decreases with σ , it is clear that L_M is maximized in B when $\alpha = \alpha_1(\sigma_0)$ and $\sigma = \sigma_0$. Thus, the threshold value of δ above which a mediation contract can be sustained, δ_M , is given by the equation $L_M(\alpha_1(\sigma_0), \sigma_0, \delta_M) = v_0$ or, equivalently,

$$(1 - \delta_M) \left(\frac{1 - \sigma_0}{2 - \sigma_0} \right) y + \delta_M \left[\frac{c}{b}y + \left(1 - \frac{c}{b} \right) v_0 \right] = v_0.$$

Solving this equation with respect to δ_M , we obtain

$$\delta_M = \frac{1}{1 + \frac{c}{b}S_M}, \quad (7)$$

where $S_M = ((y/v_0) - 1) / \left(1 - \left(\frac{1 - \sigma_0}{2 - \sigma_0} \right) (y/v_0) \right)$.

For any $\delta \geq \delta_M$ and $\sigma \geq \sigma_0$, define $\alpha_M(\sigma, \delta)$ as the value of α for which the constraint (IC_M^M) holds with equality. Since L_M decreases with α , we immediately obtain that (IC_M^M) is satisfied if and only if $\alpha \leq \alpha_M(\sigma, \delta)$. We next show that $\alpha_M(\sigma, \delta)$ increases with δ , but need not be monotone in σ . Using the Implicit Function theorem, we obtain

$$\begin{aligned} \frac{\partial}{\partial \delta} \alpha_M(\sigma, \delta) &= -\frac{\partial L_M / \partial \delta}{\partial L_M / \partial \alpha} = \frac{2}{\delta y} \left(v^* - \frac{1 - \sigma}{2 - \sigma} y \right) > 0 \text{ and} \\ \frac{\partial}{\partial \sigma} \alpha_M(\sigma, \delta) &= -\frac{\partial L_M / \partial \sigma}{\partial L_M / \partial \alpha} = \frac{2}{\delta y} \left[-(1 - \delta) \frac{y}{(2 - \sigma)^2} + \delta \frac{1 - \frac{c}{b}}{(2 - \sigma)^2} y \right]. \end{aligned}$$

Thus, $\alpha_M(\sigma, \delta)$ increases with δ , and it increases with σ if and only if $\delta > \frac{b}{2b - c} =: \hat{\delta}$. Whether $\delta_M \leq \hat{\delta}$ depends on whether

$$L_M(\alpha_1(\sigma_0), \sigma_0, \hat{\delta}) = (1 - \hat{\delta}) \left(\frac{1 - \sigma_0}{2 - \sigma_0} \right) y + \hat{\delta} \left[\frac{c}{b}y + \left(1 - \frac{c}{b} \right) v_0 \right] \geq v_0,$$

which is equivalent to

$$\frac{1}{2} + \frac{c}{b} \left(1 - \frac{\pi_0}{Y} \right) \geq \frac{v_0}{y}.$$

Hence, when this condition is satisfied, $\alpha_M(\sigma, \delta)$ decreases in σ if $\delta \in [\delta_M, \hat{\delta}]$ and increases in σ if $\delta \in [\hat{\delta}, 1)$. And when this condition is violated, $\alpha_M(\sigma, \delta)$ increases in σ for all $\delta \in [\delta_M, 1)$. $\square$

Proof of Proposition 3. As argued in the text, the intervention contract can be sustained if and only if the constraints (IC_M^I) and (IC_P^I) are satisfied.

Consider first the constraint (IC_M^I) . As shown in the proof of Proposition 2, we can write

$$v^* = \left(1 - \frac{c}{b}\right) v_0 + \left(\frac{1}{2-\sigma} + \frac{c}{b} \left(\frac{1-\sigma}{2-\sigma}\right) - \frac{1}{2}\alpha\right) y. \quad (8)$$

Thus, $v^* \geq v_0$ if and only if

$$\alpha \leq \alpha_I(\sigma) := \frac{2}{2-\sigma} - \frac{2c}{b} \left(\frac{v_0}{y} - \frac{1-\sigma}{2-\sigma}\right),$$

meaning that (IC_M^I) holds if and only if $\alpha \leq \alpha_I(\sigma)$. Since $\partial \alpha_I(\sigma) / \partial \sigma = 2(1 - \frac{c}{b}) / (2 - \sigma)^2$ and by assumption $b > c$, we obtain that $\alpha_I(\sigma)$ increases in σ .

Consider now the constraint (IC_P^I) . Recall that this constraint requires that

$$(1 - \delta) \frac{1}{2} Y + \delta \pi^* \geq \pi_0.$$

From (2) and the fact that

$$\eta^* = 1 - \frac{1 - \frac{c}{b}}{\alpha \left(1 - \frac{1}{2}\sigma\right)},$$

we obtain that

$$\pi^* = \left(\frac{1}{2}\sigma \left(\frac{1}{2-\sigma}\right) + \frac{c}{b} \left(\frac{1-\sigma}{2-\sigma}\right)\right) Y + \left(1 - \frac{c}{b}\right) \pi_0, \quad (9)$$

which is independent of α . Thus, the left-hand side of (IC_P^I) is also independent of α , and, in what follows, we denote it by $L_I(\sigma, \delta)$. Observe that

$$\frac{\partial L_I}{\partial \sigma} = \frac{\delta Y}{(2-\sigma)^2} \left(1 - \frac{c}{b}\right) > 0.$$

Hence, L_I increases in σ . Therefore, it is maximized in B when $\sigma = 1$ (which corresponds to the subset of points $\{(\alpha, \sigma) : \alpha \geq \alpha_1(\sigma), \sigma = 1\}$ in B). Moreover, note that (IC_M^I) is satisfied when $(\alpha, \sigma) = (\alpha_1(1), 1)$ since $v^*|_{(\alpha, \sigma) = (\alpha_1(1), 1)} = \frac{c}{b}y + (1 - \frac{c}{b})v_0 > v_0$ (i.e. $\alpha_1(1) \leq \alpha_I(1)$). Thus, the threshold value of δ above which the intervention contract be sustained, δ_I , is given by the equation $L_I(1, \delta_I) = \pi_0$ or, equivalently,

$$(1 - \delta_I) \frac{1}{2} Y + \delta_I \pi^*|_{\sigma=1} = \pi_0.$$

Solving this equation (using (9)), we obtain

$$\delta_I = \frac{\pi_0 - \frac{1}{2}Y}{(1 - \frac{c}{b})\pi_0}.$$

For any $\delta \geq \delta_I$ (and $\sigma \geq \sigma_0$), define $\sigma_I(\delta)$ as the value of σ for which (IC_P^I) holds with equality. Since L_I increases in σ , (IC_P^I) is satisfied if and only if $\sigma \geq \sigma_I(\delta)$. Moreover,

$$\frac{\partial}{\partial \delta} \sigma_I(\delta) = -\frac{\partial L_I / \partial \delta}{\partial L_I / \partial \sigma} = -\frac{(2 - \sigma)^2}{\delta Y (1 - \frac{c}{b})} \left(\pi^* - \frac{1}{2}Y \right).$$

By Lemma 1, $\pi^* \geq \pi_0 > \frac{1}{2}Y$. Hence, $\partial \sigma_I(\delta) / \partial \delta < 0$, implying that $\sigma_I(\delta)$ decreases in δ . Finally, it follows from the definition of δ_I and of σ_I that $\sigma_I(\delta_I) = 1$. And note that when $\delta = 1$, (IC_P^I) reduces to $\pi^* \geq \pi_0$, implying that $\sigma_I(1) = \sigma_0$. $\square$

Proof of Lemma 2. Step 1. If $v^* < v_0$, then by Lemma 1 the equilibrium is unique, and in this equilibrium all players take their outside options in all periods. So, in this case, the claim is trivial.

Step 2. Next, consider the case where $v^* \geq v_0$. We prove the result by contradiction. Suppose there exist equilibria in which, in some period(s), all players participate and the agent exerts effort. Consider one such equilibrium, arbitrarily chosen.

Step 2a. First, we claim that there cannot exist any history $h^{\mathcal{P},t}$ on the equilibrium path such that, given $h^{\mathcal{P},t}$, the manager babbles and the principal implements the project (regardless of the manager's recommendation). To see this, note that implementing the project in such circumstances would have to be incentive compatible for the principal:

$$(1 - \delta) \frac{1}{2}Y + \delta \tilde{\pi} \geq \pi_0,$$

where $\tilde{\pi} := \mathbb{E}[\pi_{t+1}(h^{\mathcal{P},t+1}) \mid h^{\mathcal{P},t}, d_t, e_t, x_t, r_t, \iota_t = 1]$ denotes the principal's expected payoff from the continuation game. However, this condition cannot be satisfied because $\tilde{\pi} \leq \pi^*$ (by Lemma 1, $\pi_{t+1}(h^{\mathcal{P},t+1}) \leq \pi^*$ for each $h^{\mathcal{P},t+1}$ on the equilibrium path) and (IC_P^I) is violated (as we are assuming that the intervention contract cannot be sustained in equilibrium).

Step 2b. Next, we claim that there cannot exist any history $h^{\mathcal{P},t}$ on the equilibrium path such that, given $h^{\mathcal{P},t}$, the implementation is contingent on the manager's recommendation ($\iota_t = 1$ if and only if $r_t = 1$), and the manager may recommend the project if $(\omega_t, s_t) = (C, 0)$. To see this, observe that recommending the project would have to be incentive compatible for the manager:

$$(1 - \delta) \frac{1 - \sigma}{2 - \sigma} y + \delta \tilde{v} \geq v_0,$$

where $\tilde{v} := \mathbb{E}[v_{t+1}(h^{\mathcal{M},t+1}) \mid h^{\mathcal{M},t}, d_t, e_t, x_t, z_t, r_t = 1]$ denotes the manager's expected payoff from the continuation game. However, this condition cannot be satisfied because $\tilde{v} \leq v^*$ (by Lemma 1,

$v_{t+1}(h^{\mathcal{M},t+1}) \leq v^*$ for each $h^{\mathcal{M},t+1}$ on the equilibrium path) and $(IC_{\mathcal{M}}^M)$ is violated (as we are assuming a mediation contract cannot be sustained in equilibrium).

Step 2c. Now, note that the two situations considered above are the only ones in which a project can get implemented if $(\omega, s) = (C, 0)$ (recall that we focus on equilibria in pure strategies). Since none of them are feasible, in equilibrium, no project with $(\omega, s) = (C, 0)$ is implemented.

Step 2d. Denote $w_\tau := \Pr[e_\tau = 1 \mid h^{\mathcal{A},\tau}]$, $p_\tau := \Pr[l_\tau = 1 \mid (\omega_\tau, s_\tau) \neq (C, 0), e_\tau = 1, h^{\mathcal{A},t}]$, and $q_\tau := \Pr[l_\tau = 1 \mid (\omega_\tau, s_\tau) = (C, 0), e_\tau = 1, h^{\mathcal{A},t}] = 0$. Then, consider the decomposition of the agent's payoff u at the beginning of period 1:

$$\begin{aligned} u &= (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} w_\tau [\Pr((\omega_\tau, s_\tau) \neq (C, 0)) p_\tau b + \Pr((\omega_\tau, s_\tau) = (C, 0)) q_\tau b - c] \\ &= (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} w_\tau [\Pr((\omega_\tau, s_\tau) \neq (C, 0)) p_\tau b - c] \\ &\leq (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} w_\tau [\Pr((\omega_\tau, s_\tau) \neq (C, 0)) b - c]. \end{aligned}$$

As $u \geq 0$ in any equilibrium, and $\Pr((\omega_\tau, s_\tau) \neq (C, 0)) b < c$ for $(\alpha, \sigma) \in B$, we must have $w_\tau = 0$ for all $\tau \geq 1$. Therefore, in the equilibrium we consider, $e_\tau = 0$ for all $\tau \geq 1$, which contradicts with the assumption that the agent may exert effort in some period(s). $\square$

Proof of Theorem 1. Fix $\delta \in [\delta_I, 1)$. Since $\delta \geq \delta_I$, we know by Proposition 3 that $B_I \neq \emptyset$. We next analyze B_M and how it compares to B_I .

Step 1. By Proposition 2, we know that $B_M \neq \emptyset$ if and only if $\delta \geq \delta_M$. From (7) in the proof of Proposition 2, we obtain that δ_M is strictly decreasing in y/v_0 , $\delta_M \rightarrow 1$ as $y/v_0 \rightarrow 1$, and $\delta_M \rightarrow 0$ as $y/v_0 \rightarrow (2 - \sigma_0)/(1 - \sigma_0)$. Since δ_M is continuous and decreasing in y/v_0 , there exists $\tau_1 \in (1, (2 - \sigma_0)/(1 - \sigma_0))$ such that $\delta_M = \delta$ if $y/v_0 = \tau_1$, and $\delta_M \leq \delta$ if and only if $y/v_0 \geq \tau_1$. Hence, $B_M = \emptyset$ if $y/v_0 < \tau_1$. This establishes part (i) of the theorem.

Step 2. We now prove part (iii) of the theorem. Consider the point $(\alpha, \sigma) = (1, \sigma_I(\delta)) \in B$. If it is also the case that $(1, \sigma_I(\delta)) \in B_M$, then by Proposition 3 and Proposition 4, we obtain that $B_I \subseteq B_M$. Now, observe that $(1, \sigma_I(\delta)) \in B_M$ if and only if $(IC_{\mathcal{M}}^M)$ is satisfied at this point, i.e. if and only if

$$(1 - \delta) \left(\frac{1 - \sigma_I}{2 - \sigma_I} \right) y + \delta v^* \Big|_{(\alpha, \sigma) = (1, \sigma_I)} \geq v_0. \quad (10)$$

Plugging (8) in this inequality and rearranging, we obtain

$$\frac{y}{v_0} \geq \frac{(2 - \sigma_I) [1 - \delta(1 - \frac{\varepsilon}{b})]}{(1 - \delta)(1 - \sigma_I) + \frac{1}{2} \delta \sigma_I + \delta \frac{\varepsilon}{b} (1 - \sigma_I)} =: \tau_2.$$

Thus, if $y/v_0 \geq \tau_2$, then $B_I \subseteq B_M$. Furthermore, observe that by definition of σ_I , $(IC_{\mathcal{P}}^I)$ holds with

equality at the point $(\alpha, \sigma) = (1, \sigma_I(\delta))$, implying that

$$(1 - \delta) \frac{1}{2} Y + \delta \pi^* \big|_{(\alpha, \sigma) = (1, \sigma_I)} = \pi_0.$$

Plugging (2) in this equation and rearranging, we obtain

$$\frac{Y}{\pi_0} = \frac{(2 - \sigma_I) [1 - \delta(1 - \frac{c}{b})]}{(1 - \delta)(1 - \sigma_I) + \frac{1}{2}\sigma_I + \delta \frac{c}{b}(1 - \sigma_I)}.$$

As $\delta < 1$, combining the above equation and the expression for τ_2 derived above, we obtain that $\tau_2 > Y/\pi_0$.

Step 3. We now argue that $\tau_2 > \tau_1$.

Step 3a. Recall that whenever $y/v_0 \geq \tau_2$, we have $(1, \sigma_I(\delta)) \in B_M$. Then, as $B_M \neq \emptyset$, $y/v_0 \geq \tau_1$. Since $y/v_0 \geq \tau_1$ for any ratio $y/v_0 \geq \tau_2$, we conclude $\tau_2 \geq \tau_1$.

Step 3b. Next, we rule out the case that $\tau_2 = \tau_1$. We prove this by contradiction. Suppose that $\tau_2 = \tau_1$, and let $y/v_0 = \tau_1 = \tau_2$. Then, by the definition of τ_1 and the definition of δ_M (see the proof of [Proposition 2](#)), we know that $\delta = \delta_M$, and

$$(1 - \delta) \left(\frac{1 - \sigma_0}{2 - \sigma_0} \right) y + \delta v^* \big|_{(\alpha, \sigma) = (\alpha_1(\sigma_0), \sigma_0)} = v_0.$$

Since $\sigma_I(\delta) > \sigma_0$ (see [Proposition 3](#)),

$$\frac{1 - \sigma_0}{2 - \sigma_0} > \frac{1 - \sigma_I(\delta)}{2 - \sigma_I(\delta)}.$$

In addition, the proof of [Proposition 2](#) has shown that

$$v^* \big|_{(\alpha, \sigma) = (\alpha_1(\sigma_0), \sigma_0)} > v^* \big|_{(\alpha, \sigma) = (1, \sigma_I(\delta))}.$$

Combining these two facts, we have

$$(1 - \delta) \frac{1 - \sigma_I(\delta)}{2 - \sigma_I(\delta)} y + \delta v^* \big|_{(\alpha, \sigma) = (1, \sigma_I(\delta))} < v_0,$$

implying that (10) fails and $y/v_0 < \tau_2$. This is a contradiction.

Step 4. Finally, we prove part (ii) of the theorem. That is, we show that if $y/v_0 \in [\tau_1, \tau_2)$, then $B_M \setminus B_I \neq \emptyset$ and $B_I \setminus B_M \neq \emptyset$.

Step 4a. Since $y/v_0 \geq \tau_1$, then $B_M \neq \emptyset$, which implies that $(\alpha_1(\sigma_0), \sigma_0) \in B_M$. However, because $\sigma_I(\delta) > \sigma_0$, we obtain that $(\alpha_1(\sigma_0), \sigma_0) \notin B_I$ and, therefore, $B_M \setminus B_I \neq \emptyset$.

Step 4b. To show that $B_I \setminus B_M \neq \emptyset$, we consider two possible cases separately: (i) the case where the constraint $\alpha \leq \alpha_I(\sigma)$ is satisfied at $(\alpha, \sigma) = (1, \sigma_I(\delta))$ and (ii) the case when it is not. Regarding

the first case, since $\alpha \leq \alpha_I(\sigma)$ is satisfied at $(\alpha, \sigma) = (1, \sigma_I(\delta))$, then $(1, \sigma_I(\delta)) \in B_I$; and because $y/v_0 < \tau_2$, $(1, \sigma_I(\delta)) \notin B_M$. Thus, $B_I \setminus B_M \neq \emptyset$.

Consider now the second case where $\alpha \leq \alpha_I(\sigma)$ is not satisfied at $(\alpha, \sigma) = (1, \sigma_I(\delta))$. Because $\alpha \leq \alpha_I(\sigma)$ is violated at $(\alpha, \sigma) = (1, \sigma_I(\delta))$, this point does not satisfy $(IC_{\mathcal{M}}^I)$, implying that $(1, \sigma_I(\delta)) \notin B_I$. However, the point $(\alpha, \sigma) = (\alpha_I(\sigma_I(\delta)), \sigma_I(\delta))$ where $(IC_{\mathcal{M}}^I)$ just holds, satisfies both $(IC_{\mathcal{M}}^I)$ and $(IC_{\mathcal{P}}^I)$ and, therefore, is in B_I . Now, since $(IC_{\mathcal{M}}^I)$ holds with equality at $(\alpha_I(\sigma_I(\delta)), \sigma_I(\delta))$, we know that $v^* = v_0$ at this point. Because $v^* = v_0$ and $\frac{1-\sigma_I}{2-\sigma_I}y \leq \frac{1}{2}y < v_0$, we obtain that $(IC_{\mathcal{M}}^M)$ is necessarily violated at $(\alpha_I(\sigma_I(\delta)), \sigma_I(\delta))$, and $(\alpha_I(\sigma_I(\delta)), \sigma_I(\delta)) \notin B_M$. Since $(\alpha_I(\sigma_I(\delta)), \sigma_I(\delta)) \in B_I$ and $(\alpha_I(\sigma_I(\delta)), \sigma_I(\delta)) \notin B_M$, $B_I \setminus B_M \neq \emptyset$. $\square$

Proof of Lemma 3. That δ_M is decreasing in y/v_0 follows directly from (7) and the observation that S_M is increasing in y/v_0 .

Next, note that, by definition, at α_M $(IC_{\mathcal{M}}^M)$ binds. Thus, we have

$$\tilde{L}_M := (1 - \delta) \left(\frac{1 - \sigma}{2 - \sigma} \right) \frac{y}{v_0} + \delta \frac{v^*}{v_0} \Big|_{\alpha = \alpha_M} \equiv 1.$$

Also, from the expression for v^* as given in the proof of Proposition 2), it readily follows that $\partial(v^*/v_0)/\partial(y/v_0) > 0$. Thus, using the Implicit Function theorem, we obtain

$$\frac{\partial \alpha_M}{\partial(y/v_0)} = - \frac{\partial \tilde{L}_M / \partial(y/v_0)}{\partial \tilde{L}_M / \partial \alpha} = \frac{2}{\delta y} \left((1 - \delta) \left(\frac{1 - \sigma}{2 - \sigma} \right) + \delta \left(\frac{\partial(v^*/v_0)}{\partial(y/v_0)} \right) \right) > 0.$$

$\square$

Proof of Corollary 1. Suppose that $\{z_t\}_{t \geq 1}$ is public and fix some $(\alpha, \sigma) \in B \setminus B_{\mathcal{I}}$. Consider an arbitrary equilibrium where at least in one period all players participate and the agent exerts effort.

Step 1. We first prove that in this equilibrium, there exists a period t and a realization of (x_t, z_t) such that the manager recommends the project even if $(\omega_t, s_t) = (C, 0)$.

Suppose not. Then, from Step 2d in the proof of Lemma 2, we know that the agent's payoff u at the beginning of period 1 is

$$u \leq (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} w_{\tau} [\Pr((\omega_{\tau}, s_{\tau}) \neq (C, 0))b - c] < 0.$$

But this is a contradiction as in any equilibrium, $u \geq 0$.

Step 2. As in equilibrium, the manager recommends the project when $(\omega_t, s_t) = (C, 0)$, it is optimal for him to also recommend the project when $(\omega_t, s_t) = (D, 0)$. Then, in the principal's incentive constraint

$$(1 - \delta) \Pr(V_{\mathcal{P}} = Y \mid r = 1, x_t, z_t) Y + \delta \pi^* \geq \pi_0,$$

we have $\Pr(V_P = Y \mid r = 1, x_t, z_t) \leq \frac{1}{2}$. Therefore, (IC_P^I) must be satisfied. Furthermore, as we assume that all players participate in at least one period, (IC_M^I) must hold as well. But this implies $(\alpha, \sigma) \in B_I$, and contradicts the fact that $(\alpha, \sigma) \in B \setminus B_I$. $\square$

References

- [1] Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1990. "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica*, Vol. 58 (5); pp. 1041–1063.
- [2] Adhvaryu, A., A. Nyshadham, and J. Tamayo. 2023. "Managerial quality and productivity dynamics." *The Review of Economic Studies*, Vol. 90; pp. 1569–1607.
- [3] Adhvaryu, A., N. Kala, and A. Nyshadham. 2022. "Management and shocks to worker productivity." *Journal of Political Economy*, Vol. 130; pp. 1–47.
- [4] Aghion, Philippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, Vol. 105; pp. 1–29.
- [5] Aghion, Philippe, John Van Reenen, and Luigi Zingales. 2013. "Innovation and Institutional Ownership." *American Economic Review*, Vol. 103 (1); pp. 277–304.
- [6] Alonso, Ricardo, Wouter Dessein, and Niko Matouschek. 2008. "When Does Coordination Require Centralization?" *American Economic Review*, Vol. 98 (1); pp. 145–79.
- [7] Alonso, Ricardo, and Niko Matouschek. 2007a. "Optimal Delegation." *Review of Economic Studies*, Vol. 75, pp. 259–293.
- [8] Alonso, Ricardo, and Niko Matouschek. 2007b. "Relational Delegation" *RAND Journal of Economics*, Vol. 38 (4); pp. 1070–1089.
- [9] Bianchi, Nicola, and Michela Giorcelli. 2022. "The dynamics and spillovers of management interventions: Evidence from the training within industry program." *Journal of Political Economy*, Vol. 130; pp. 1630–1675.
- [10] Athey, Susan, and John Roberts. 2001. "Organizational Design: Decision Rights and Incentive Contracts." *American Economic Review*, Vol. 91 (2); pp. 200–205.
- [11] Baker, George, Robert Gibbons, and Kevin J. Murphy. 1999. "Informal authority in organizations." *Journal of Law, Economics & Organization*, Vol. 15; pp. 56–73.
- [12] Baker, George, Robert Gibbons, and Kevin J. Murphy. 2002. "Relational contracts and the theory of the firm." *Quarterly Journal of Economics*, Vol. 117; pp. 39–84.
- [13] Bensaou, Ben. "Built to Innovate: Essential Practices to Wire Innovation into Your Company's DNA." McGraw Hill, 2021.

- [14] Bertrand, M., and A. Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *The Quarterly Journal of Economics*, Vol. 118; pp. 1169–1208.
- [15] Board, Simon. 2011. "Relational Contracts and the Value of Loyalty." *American Economics Review*, Vol. 101; pp. 3349–3367.
- [16] Burgelman Robert A. 1983a. "A model of the interaction of strategic behavior, corporate context, and the concept of strategy." *The Academy of Management Review*, Vol. 8 (1); pp. 61–70.
- [17] Burgelman, Robert A. 1983b. "Corporate entrepreneurship and strategic management: Insights from a process study." *Management Science*, Vol. 29 (12); pp. 1349–1364.
- [18] Burgelman, Robert A. 1991. "Intraorganizational ecology of strategy making and organizational adaptation: Theory and field research." *Organization Science*, Vol. 2 (3); pp. 239–262.
- [19] Burgelman, Robert A. 1994. "Fading memories: A process theory of strategic business exit in dynamic environments." *Administrative Science Quarterly*, Vol. 39 (1); pp. 24–56.
- [20] Burguet, Roberto, and Yeon-Koo Che. 2004. "Competitive Procurement with Corruption." *The RAND Journal of Economics*, Vol. 35 (1); pp. 50–68.
- [21] Chandler Jr, Alfred D. "The Visible Hand: The Managerial Revolution in American Business" Harvard university press, 1977.
- [22] Chassang, Sylvain. 2010. "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts." *American Economic Review*, Vol. 100 (1); pp. 448–65.
- [23] Dalton, Melville. "Men Who Manage: Fusions of Feeling and Theory in Administration." John Wiley & Sons, 1959.
- [24] Dessein, W. 2002. "Authority and Communication in Organizations." *Review of Economic Studies*, Vol. 69; pp. 811–838.
- [25] Dessein, W., and T. Santos. 2021. "Managerial Style and Attention." *American Economic Journal: Microeconomics*, Vol. 13; pp. 372–403.
- [26] Dessein, W., and L. Garicano. 2025. "La Cordata: Loyalty in Political Tournaments." *Working Paper*.
- [27] Egorov, Georgy, and Konstantin Sonin. 2001. "Dictators and Their Viziers: Endogenizing the Loyalty–Competence Trade-off." *Journal of the European Economic Association*, Vol. 9 (5); pp. 903–930.
- [28] Fairburn, James A., and James M. Malcomson. 2001. "Performance, Promotion, and the Peter Principle." *Review of Economic Studies*, Vol. 68 (1); pp. 45–66.

- [29] Farson, Richard, and Ralph Keyes. "Whoever Makes the Most Mistakes Wins: The Paradox of Innovation." The Free Press: New York, NY. 2002.
- [30] Fenizia, Alessandra. 2022. "Managers and productivity in the public sector." *Econometrica*, Vol. 90 (3); pp. 1063–1084.
- [31] Foss, N. J. and P. G Klein. "Why Managers Matter: The Perils of the Bossless Company." New York: PublicAffairs, 2022.
- [32] Fong, Yuk-Fai, and Jin Li. 2017. "Information Revelation in Relational Contracts." *Review of Economic Studies*, Vol. 84 (1); pp. 277–299.
- [33] Friebe G., Heinz M, and N. Zubanov. 2022. "Middle Managers, Personnel Turnover, and Performance: A Long-Term Field Experiment in a Retail Chain." *Management Science*, Vol. 68; pp. 211–229
- [34] Friebe, Guido and Michael Raith. 2022. "Talent Management: The Role of Bosses." *SSRN Working Paper*.
- [35] Fuchs, William. 2007. "Contracting with Repeated Moral Hazard and Private Evaluations." *American Economic Review*, Vol. 97 (4); pp. 1432–1448.
- [36] Gibbons, Robert, Niko Matouschek, and John Roberts. "Decisions in Organizations." *The Handbook of Organizational Economics*, edited by Robert Gibbons and John Roberts, Princeton: Princeton University Press, 2012; pp. 373–431.
- [37] Hirschman, A. O. "Exit, voice, and loyalty: Responses to decline in firms, organizations, and states." Harvard university press, 1970.
- [38] Hoffman, Mitchell, and Steven Tadelis. 2021. "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis." *Journal of Political Economy*, Vol. 129 (1); pp. 243–285.
- [39] Holmström, Bengt. "On Incentives and Control in Organizations." Ph.D. Dissertation, Graduate School of Business, Stanford University, 1977.
- [40] Holmström, Bengt. "On the Theory of Delegation." In M. Boyer and R. Kihlstrom, eds., *Bayesian Models in Economic Theory*. New York: North-Holland, 1984.
- [41] House, C., and R. Price. "The HP Phenomenon: Innovation and Business Transformation." 1st ed. Stanford University Press, Stanford, CA. 2009.
- [42] Isaacson, W. "Steve Jobs." Simon & Schuster, 2011.
- [43] Jaser, Zahira. 2021. "The Real Value of Middle Managers." *Harvard Business Review*, Vol. 99 (3); pp. 124–133.

- [44] Kofman, Fred and Jacques Lawarree. 1993. "Collusion in Hierarchical Agency." *Econometrica*, Vol. 61, pp. 629–656.
- [45] Kolotilin, Anton, and Hongyi Li. 2021. "Relational communication." *Theoretical Economics*, Vol. 16 (4); pp. 1391–1430.
- [46] Kuvalekar, Aditya, Elliot Lipnowski, and João Ramos. 2022. "The Real Value of Middle Managers." *Journal of Economic Theory*, Vol. 203; 105467.
- [47] Kvaløy, Ola, and Trond E. Olsen. 2009. "Endogenous Verifiability and Relational Contracting." *American Economic Review*, Vol. 99 (5); pp. 2193–2208.
- [48] Levitt, S. D., and C. M. Snyder. 1997. "Is no News Bad News? Information Transmission and the Role of "Early Warning" in the Principal-Agent Model." *RAND Journal of Economics*, Vol. 28; pp. 641–661.
- [49] Li, Jin, Matouschek, Niko, and Michael Powell. 2017. "Power Dynamics in Organizations." *American Economic Journal: Microeconomics*, Vol 9; pp. 217–41.
- [50] Lipsky, Michael. "Street-Level Bureaucracy: Dilemmas of the Individual in Public Services." Russell Sage Foundation, " 30th Anniversary, Expanded ed. Ed. 2010.
- [51] Lipnowski, Elliot, and João Ramos. 2020. "Repeated Delegation." *Journal of Economic Theory*, Vol. 188; 105040.
- [52] Lo D, Brahm, F., Dessein, W., and C. Minami. 2022. "Managing with Style? Micro-evidence on the Allocation of managerial Attention." *Management Science*, Vol 68; pp. 8261–8285.
- [53] Metcalfe, Robert D., Alexandre B. Sollaci, and Chad Syverson. 2023. "Managers and productivity in retail." NBER Working Paper 31192.
- [54] Manso, Gustavo. 2011. "Motivating Innovation." *Journal of Finance*, Vol. 66; pp. 1823–1860.
- [55] Mintzberg, H. "Mintzberg on Management: inside our strange world of organizations." The Free Press, 1989.
- [56] Mookherjee, Dilip. "Incentives in Hierarchies." *The Handbook of Organizational Economics*, edited by Robert Gibbons and John Roberts, Princeton: Princeton University Press, 2013, pp. 764–798.
- [57] Nguyen, Kieu-Trang. 2023. "Trust and Innovation within the Firm: Evidence from Matched CEO-Firm Data." *Working Paper*.
- [58] Olsen, Trond E. and Gaute Torsvik (1998). "Collusion and Renegotiation in Hierarchies: A Case of Beneficial Corruption." *International Economic Review*, Vol. 39, pp. 413–438.

- [59] Orlov, Dmitry. 2022. "Frequent Monitoring in Dynamic Contracts." *Journal of Economic Theory*, Vol. 206; 105550.
- [60] Osterman, P. "The truth about middle managers: Who they are, how they work, why they matter." Harvard Business Review Press, 2009.
- [61] Prendergast, Canice. 2002. "Consumers and Agency Problems." *Economics Journal*, Vol. 112 (March); pp. C34–C51.
- [62] Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency." *Journal of Political Economy*, Vol. 111 (5); pp. 929–958.
- [63] Rantakari, Heikki. 2008. "Governing Adaptation." *Review of Economic Studies*, Vol. 75 (4); pp. 1257–1285.
- [64] Rantakari, Heikki. 2012. "Employee Initiative and Managerial Control." *American Economic Journal: Microeconomics*, Vol. 4 (3); pp. 171–211.
- [65] Ray, Debraj. 2002. "The Time Structure of Self-Enforcing Agreements." *Econometrica*, Vol. 70; pp. 547–582.
- [66] Roberts, John, and Kathryn L. Shaw. 2022. "Managers and the management of organizations." NBER Working Paper No. 30730.
- [67] Rotemberg, Julio J., and Garth Saloner. 1993. "Leadership Styles and Incentives." *Management Science*, Vol. 39 (11); pp. 1299–1318.
- [68] Rotemberg, Julio J., and Garth Saloner. 2000. "Visionaries, Managers, and Strategic Direction." *RAND Journal of Economics*, Vol. 31 (4); pp. 693–716.
- [69] Tirole, Jean. 1986. "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." *Journal of Law, Economics, and Organization*, Vol. 2 (2); pp. 181–214.
- [70] Troya-Martinez, Marta, and Liam Wren-Lewis. 2023. "Managing Relational Contracts." *Journal of the European Economic Association*, Vol. 21 (3); pp. 941–986.
- [71] Turco, Catherine J. "The Conversational Firm: Rethinking Bureaucracy in the Age of Social Media." Columbia University Press, 2016.
- [72] Vong, Allen. 2025. "Mediation and worker performance." *Working Paper*.
- [73] Wong, Yu Fu. 2023. "Dynamic Monitoring Design." *Working Paper*.
- [74] Zakharov, Alexei V. 2016. "The Loyalty-Competence Trade-Off in Dictatorships and Outside Options for Subordinates." *Journal of Politics*, Vol. 78 (2); pp. 457–466.

Online Appendix: “Managing Loyalty in Hierarchical Organizations”

In this appendix we present the proof of **Proposition 4**. To show this result, we first present several useful definitions and observations, then prove the proposition as a corollary.

Denote the payoff profile from the 0-lenient action as $(u_{l0}, v_{l0}, \pi_{l0})$ and that from the 1-lenient action as $(u_{l1}, v_{l1}, \pi_{l1})$. Note that these two profiles can be written as

$$u_{l0} = \frac{1}{2}\sigma b - c, \quad v_{l0} = \frac{1}{2}\sigma y + \left(1 - \frac{1}{2}\sigma\right)v_0, \quad \pi_{l0} = \frac{1}{2}\sigma Y + \left(1 - \frac{1}{2}\sigma\right)\pi_0,$$

and

$$\begin{aligned} u_{l1} &= \left[1 - \alpha\left(1 - \frac{1}{2}\sigma\right)\right]b - c, \quad v_{l1} = \frac{1}{2}\sigma y + (1 - \alpha)\left(1 - \frac{1}{2}\sigma\right)y + \alpha\left(1 - \frac{1}{2}\sigma\right)v_0, \\ \pi_{l1} &= \frac{1}{2}\sigma Y + \frac{1}{2}(1 - \alpha)(1 - \sigma)Y + \alpha\left(1 - \frac{1}{2}\sigma\right)\pi_0. \end{aligned}$$

The payoff profile from a β -lenient action is given by

$$(u_{l\beta}, v_{l\beta}, \pi_{l\beta}) := \beta(u_{l1}, v_{l1}, \pi_{l1}) + (1 - \beta)(u_{l0}, v_{l0}, \pi_{l0}).$$

Since $u_{l1} > u_{l0}$, $v_{l1} > v_{l0}$, and $\pi_{l1} < \pi_{l0}$, we know that with respect to the lenience level β , both $u_{l\beta}$ and $v_{l\beta}$ are strictly increasing, but $\pi_{l\beta}$ is strictly decreasing.

Define

$$\beta_* := \min\{\beta \in [0, 1] : u_{l\beta} \geq 0\}.$$

That is, β_* is the lowest lenience level that brings to the agent a zero payoff. Similarly, define

$$\beta^* := \max\{\beta \in [0, 1] : \pi_{l\beta} \geq \pi_0\}.$$

That is, β^* is the highest lenience level that brings to the principal a payoff of at least π_0 . Denote the payoff profile from the β_* -lenient action as $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ and that from the β^* -lenient action as $(u_{l\beta^*}, v_{l\beta^*}, \pi_{l\beta^*})$. We observe the following about these actions and their payoffs (**Figure 6**). Note that in this lemma, the definitions of $\alpha_0(\sigma)$, $\alpha_1(\sigma)$, and σ_0 remain the same as those given in **Proposition 1**.

Lemma 4. Suppose $\alpha < \alpha_1(\sigma)$ and $\sigma \geq \sigma_0$. The following holds:

- (i) If $\alpha < \alpha_0(\sigma)$ and $\sigma > \sigma_0$, then $0 < \beta_* < \beta^* < 1$.
- (ii) If $\alpha \geq \alpha_0(\sigma)$ and $\sigma > \sigma_0$, then $0 < \beta_* < 1$ and $\beta^* = 1$.
- (iii) If $\sigma = \sigma_0$, then $\beta_* = \beta^* \in (0, 1)$.

(iv) $v_{l\beta} > v_0$ for any $\beta \in [0, 1]$.

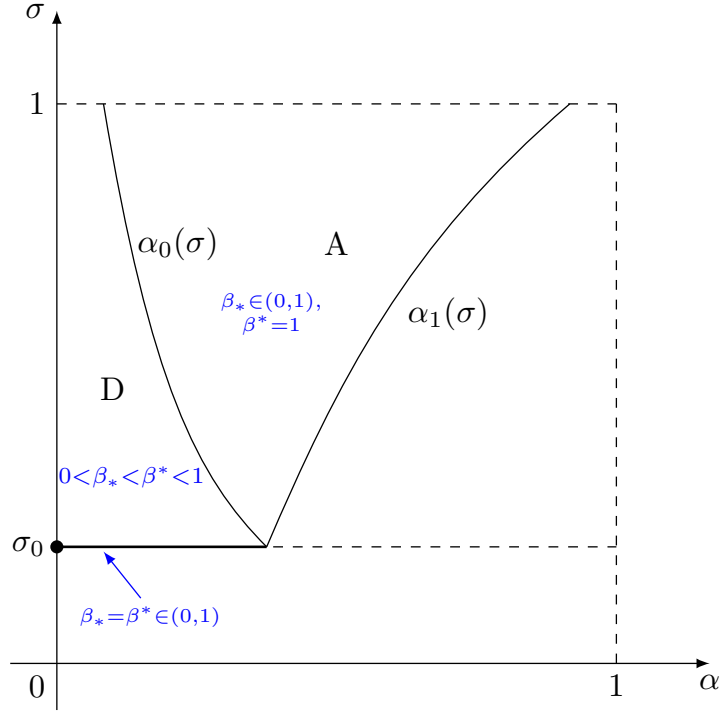


Figure 6: Illustration of Lemma 4

Proof. First, we prove the results concerning only β_* . Given the β_* -lenient action, any project with $s = 1$ is implemented with probability one, any project with $(\omega, s) = (D, 0)$ is implemented with probability β_* , and any project with $(\omega, s) = (C, 0)$ is never implemented. As the agent gains at least zero from the β_* -lenient action, we have

$$[\Pr(s = 1) + \Pr(\omega = D, s = 0)\beta_*]b = \left[\frac{1}{2}\sigma + (1 - \alpha)(1 - \frac{1}{2}\sigma)\beta_* \right] b \geq c.$$

From this inequality we can solve that

$$\beta_* = \frac{1}{2 - \sigma} \left[\frac{2c}{b} - \sigma \right], \quad (11)$$

which is strictly positive as $2c/b > 1 \geq \sigma$.

Next, consider β^* . Proposition 1 shows that the 1-lenient action is a Nash equilibrium of the stage game if and only if $\alpha \geq \alpha_0(\sigma)$. Thus, if $\alpha \geq \alpha_0(\sigma)$, then $\beta^* = 1$; and if $\alpha < \alpha_0(\sigma)$, then $\beta^* < 1$.

To compare β_* and β^* , we first write out principal's payoff from a β -lenient action:

$$\begin{aligned} \pi_{l\beta} &= \beta\pi_{l1} + (1 - \beta)\pi_{l0} \\ &= \frac{1}{2}\sigma Y + \frac{1}{2}(1 - \alpha)(1 - \sigma)\beta Y + (\alpha\beta + 1 - \beta) \left(1 - \frac{1}{2}\sigma \right) \pi_0. \end{aligned} \quad (12)$$

Because $\pi_{l1} < \pi_{l0}$, $\pi_{l\beta}$ is strictly decreasing in β . Further, from (11) and (12), it is straightforward to show that $\pi_{l\beta_*} = \pi_0$ when $\sigma = \sigma_0 := 2c(2\pi_0 - Y)/[2c(\pi_0 - Y) + Yb]$ and $\pi_{l\beta_*} > \pi_0$ when $\sigma > \sigma_0$. Since the β^* -lenient action gives the principal a payoff at least equal to π_0 , we conclude that when $\sigma = \sigma_0$, $\beta^* = \beta_*$; and when $\sigma > \sigma_0$, $\beta^* > \beta_*$. In addition, because $\alpha_0(\sigma_0) = \alpha_1(\sigma_0)$, we know that when $\sigma = \sigma_0$ and $\alpha < \alpha_1(\sigma_0)$, $\beta_* = \beta^* < 1$.

Finally, note that both the 0-lenient and the 1-lenient actions yield the manager a payoff strictly greater than v_0 . Hence, as a convex combinations of these actions, $v_{l\beta} > v_0$ for any $\beta \in [0, 1]$. $\square$

The following lemma establishes $\pi_{l\beta_*}$ as an upper bound of the principal's equilibrium payoff.

Lemma 5. *Suppose $\alpha < \alpha_1(\sigma)$ and $\sigma \geq \sigma_0$. Then, the principal's payoff is weakly smaller than $\pi_{l\beta_*}$ in any equilibrium.*

Proof. Consider an arbitrary equilibrium. Denote the principal's equilibrium payoff as π , and the agent's equilibrium payoff as u . In the equilibrium, for each $\tau \geq 1$, let $p_\tau := \Pr[\iota_\tau = 1 \mid s_\tau = 1, e_\tau = 1]$ and $q_\tau := \Pr[\iota_\tau = 1 \mid s_\tau = 0, e_\tau = 1]$ denote the implementation probability conditional on $e_\tau = 1$ and s_τ . Notice that these probabilities are evaluated at the beginning of the repeated game, where the principal and the agent share the same information. In addition, we denote $w_\tau := \Pr[e_\tau = 1]$ as the probability that the agent exerts effort at period τ .

According to Lemma 4, $\beta_* > 0$, implying $u_{l\beta_*} = 0$. We can express u and $u_{l\beta_*}$ as:

$$u = (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} w_\tau \left[\frac{1}{2} \sigma p_\tau b + \left(1 - \frac{1}{2} \sigma \right) q_\tau b - c \right],$$

$$u_{l\beta_*} = (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} w_\tau \left[\frac{1}{2} \sigma b + \left(1 - \frac{1}{2} \sigma \right) (1 - \alpha) \beta_* b - c \right].$$

Since $u \geq 0 = u_{l\beta_*}$, $b > 0$, and $p_\tau \leq 1$, we obtain

$$\sum_{\tau \geq 1} \delta^{\tau-1} w_\tau q_\tau \geq \sum_{\tau \geq 1} \delta^{\tau-1} w_\tau (1 - \alpha) \beta_*. \quad (13)$$

Then, we have

$$\begin{aligned} \pi &= \pi_0 + (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} w_\tau \left[\frac{1}{2} \sigma p_\tau (Y - \pi_0) + \left(1 - \frac{1}{2} \sigma \right) q_\tau \left(\frac{1 - \sigma}{2 - \sigma} Y - \pi_0 \right) \right] \\ &\leq \pi_0 + (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} w_\tau \left[\frac{1}{2} \sigma (Y - \pi_0) + \left(1 - \frac{1}{2} \sigma \right) (1 - \alpha) \beta_* \left(\frac{1 - \sigma}{2 - \sigma} Y - \pi_0 \right) \right]. \end{aligned} \quad (14)$$

where the inequality follows from (13), and the facts that $\frac{1-\sigma}{2-\sigma} Y < \pi_0$, $Y - \pi_0 > 0$, and $p_\tau \leq 1$.

Note that $\pi_{l\beta_*}$ can be written as

$$\pi_{l\beta_*} = \pi_0 + (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} \left[\frac{1}{2} \sigma (Y - \pi_0) + \left(1 - \frac{1}{2} \sigma \right) (1 - \alpha) \beta_* \left(\frac{1 - \sigma}{2 - \sigma} Y - \pi_0 \right) \right].$$

Since $\pi_{l\beta_*} \geq \pi_{l\beta^*} \geq \pi_0$,

$$\frac{1}{2}\sigma(Y - \pi_0) + \left(1 - \frac{1}{2}\sigma\right)(1 - \alpha)\beta_* \left(\frac{1 - \sigma}{2 - \sigma}Y - \pi_0\right) \geq 0.$$

Then, as $w_\tau \leq 1$ in (14), we obtain $\pi \leq \pi_{l\beta_*}$. $\square$

Lemma 5 implies that in regions A and D, an equilibrium is automatically optimal for the principal if her payoff is $\pi_{l\beta_*}$. Recall that $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ is induced by the β_* -lenient action. The next lemma shows that when the players are sufficiently patient, $\pi_{l\beta_*}$ can be sustained as an equilibrium payoff by repeating the β_* -lenient action; and if the players are not sufficiently patient, $\pi_{l\beta_*}$ cannot be sustained by any stationary contract.

Lemma 6. *Suppose $\alpha < \alpha_1(\sigma)$ and $\sigma \geq \sigma_0$. There exists $\bar{\delta} \in (0, 1)$ (which depends on α and σ) such that if $\delta \geq \bar{\delta}$, the β_* -lenient action is self-enforcing. Furthermore, if $\delta < \bar{\delta}$, $\pi_{l\beta_*}$ cannot be sustained as an equilibrium payoff by any stationary contract.*

Proof. Suppose $\sigma \geq \sigma_0$. We first consider the case when $\alpha_0(\sigma) \leq \alpha < \alpha_1(\sigma)$. Let the players coordinate on the public randomizing device x_t : if $x_t \in [0, \beta_*]$, then the manager sends $r_t = 1$ if and only if $(\omega_t, s_t) \neq (C, 0)$, and if $x_t \in (\beta_*, 1]$, the manager sends $r_t = 1$ if and only if $s_t = 1$. The agent exerts effort, and the principal follows the manager's recommendation. If any player is publicly known to have deviated, the repeated game is terminated.

By the definition of β_* and **Lemma 4**, $u_{l\beta_*} \geq 0$, $v_{l\beta_*} > v_0$, and $\pi_{l\beta_*} \geq \pi_0$. Therefore, their participation constraints are satisfied. To enforce the β_* -lenient action, there are four incentive constraints:

$$(1 - \delta)[-c + \Pr(s = 1)b + \Pr(\omega = D, s = 0)\beta_*b] + \delta u_{l\beta_*} \geq 0, \quad (IC_{\mathcal{A}}^{\beta_*})$$

$$(1 - \delta) \Pr(V_{\mathcal{P}} = Y \mid r = 1, x \leq \beta_*)Y + \delta \pi_{l\beta_*} \geq \pi_0, \quad (IC_{\mathcal{P},1}^{\beta_*})$$

$$(1 - \delta) \Pr(V_{\mathcal{P}} = Y \mid r = 1, x > \beta_*)Y + \delta \pi_{l\beta_*} \geq \pi_0, \text{ and} \quad (IC_{\mathcal{P},2}^{\beta_*})$$

$$\begin{aligned} (1 - \delta)v_0 + \delta v_{l\beta_*} &\geq (1 - \delta)y + \delta [\Pr(V_{\mathcal{P}} = Y \mid s = 0)v_{l\beta_*} + \Pr(V_{\mathcal{P}} = 0 \mid s = 0)v_0] \quad (IC_{\mathcal{M}}^{\beta_*}) \\ &\geq (1 - \delta)y + \delta \left[\frac{1 - \sigma}{2 - \sigma}v_{l\beta_*} + \frac{1}{2 - \sigma}v_0 \right]. \end{aligned}$$

$(IC_{\mathcal{A}}^{\beta_*})$ indicates that the agent is willing to exert effort. $(IC_{\mathcal{P},1}^{\beta_*})$ and $(IC_{\mathcal{P},2}^{\beta_*})$ indicate that the principal is willing to follow the manager's recommendation. Finally, $(IC_{\mathcal{M}}^{\beta_*})$ indicates that upon $x > \beta_*$, the manager is willing to send $r = 0$ when $(\omega, s) = (D, 0)$.

Observe that $(IC_{\mathcal{A}}^{\beta_*})$ is equivalent to $u_{l\beta_*} \geq 0$. Hence, it is satisfied. From $(IC_{\mathcal{P}})$ and $\alpha \geq \alpha_0(\sigma)$ we know that $\Pr(V_{\mathcal{P}} = Y \mid r = 1, x \leq \beta_*)Y \geq \pi_0$. Then, because $\pi_{l\beta_*} \geq \pi_0$, $(IC_{\mathcal{P},1}^{\beta_*})$ is satisfied. And since $\Pr(V_{\mathcal{P}} = Y \mid r = 1, x \leq \beta_*)Y \leq \Pr(V_{\mathcal{P}} = Y \mid r = 1, x > \beta_*)Y$, $(IC_{\mathcal{P},2}^{\beta_*})$ is also satisfied. For the manager's incentive constraint, observe that the difference between the left-hand side and the right-hand side is continuous and increasing in δ , which is strictly positive when $\delta = 1$ (recall

that $v_{l\beta_*} > v_0$) and strictly negative when $\delta = 0$. Hence, there exists $\bar{\delta} \in (0, 1)$ such that $(IC_M^{\beta_*})$ holds if and only if $\delta \geq \bar{\delta}$. This completes our proof for the case when $\alpha_0(\sigma) \leq \alpha < \alpha_1(\sigma)$.

Next, we consider the case when $\alpha < \alpha_0(\sigma)$. In this case, $\Pr(V_P = Y \mid r = 1, x \leq \beta_*)Y < \pi_0$, so $(IC_{P,1}^{\beta_*})$ may be violated. However, we can resolve this issue with the private randomizing device z_t . Instead of x_t , let the players coordinate on z_t : if $z_t \in [0, \beta_*]$, then $r_t = 1 \iff (\omega_t, s_t) \neq (C, 0)$; if $z_t \in (\beta_*, 1]$, then $r_t = 1 \iff s_t = 1$; and if the manager is known to have deviated by the agent, they leave the game forever. As the principal does not observe z_t , her incentive constraint is simply

$$(1 - \delta) \Pr(V_P = Y \mid r = 1, \beta_*)Y + \delta \pi_{l\beta_*} \geq \pi_0.$$

Then, since $\pi_{l\beta_*} \geq \pi_0$ and $\pi_{l\beta_*} = \Pr(r = 1) \Pr(V_P = Y \mid r = 1, \beta_*)Y + \Pr(r = 0)\pi_0$, we have $\Pr(V_P = Y \mid r = 1, \beta_*)Y \geq \pi_0$. Together with $\pi_{l\beta_*} \geq \pi_0$, this implies that the above incentive constraint is satisfied. With this fact established, the argument in the previous paragraph follows.

Finally, suppose $\delta < \bar{\delta}$. If $\pi_{l\beta_*}$ is sustained as an equilibrium payoff by some stationary contract, it is implied that the β_* -lenient action is self-enforcing. However, when $\delta < \bar{\delta}$, the β_* -lenient action cannot be self-enforcing because $(IC_M^{\beta_*})$ is violated. $\square$

Lemma 6 shows that there exists $\bar{\delta}$ such that $\pi_{l\beta_*}$ can be sustained as an equilibrium payoff by stationary contracts if and only if $\delta \geq \bar{\delta}$. The proof of the lemma reveals two typical methods to enforce the β_* -lenient action: one through coordination on x_t and the other on z_t . These two methods can be used to enforce actions at any lenience level. Thus, it is worthwhile to formally define them.

We refer to the following as *enforcing a β -action through coordination on x_t* : If $x_t \in [0, \beta]$, then the manager sends $r_t = 1$ if and only if $(\omega_t, s_t) \neq (C, 0)$, and if $x_t \in (\beta, 1]$, then the manager sends $r_t = 1$ if and only if $s_t = 1$. The agent exerts effort, and the principal follows the manager's recommendation. If any player is publicly known to have deviated, the repeated game is terminated.

Similarly, we refer to the following as *enforcing a β -action through coordination on z_t* : If $z_t \in [0, \beta]$, then the manager sends $r_t = 1$ if and only if $(\omega_t, s_t) \neq (C, 0)$, and if $z_t \in (\beta, 1]$, then the manager sends $r_t = 1$ if and only if $s_t = 1$. The agent exerts effort, and the principal follows the manager's recommendation. If any player is publicly known to have deviated, the repeated game is terminated. Additionally, if the agent observes the manager's deviation, both the manager and the agent leave the game forever.

Using these methods, the next lemma shows that the β^* -lenient action can also be self-enforcing as long as the players are sufficiently patient. For $\sigma > \sigma_0$, the required patience level is strictly lower than $\bar{\delta}$.

Lemma 7. Suppose $\alpha < \alpha_1(\sigma)$ and $\sigma > \sigma_0$. There exists $\underline{\delta}_1 \in (0, \bar{\delta})$ (which depends on α and σ) such that if $\delta \geq \underline{\delta}_1$, the β^* -lenient action is self-enforcing.

Proof. First, consider the case when $\alpha_0(\sigma) \leq \alpha < \alpha_1(\sigma)$. In this case, we know $\beta^* = 1$ by **Lemma 4**. Then, according to **Proposition 1**, the β^* -lenient action is automatically self-enforcing. We do not

even need either x_t or z_t to enforce this action. This completes our proof for the case when $\alpha_0 \leq \alpha < \alpha_1(\sigma)$.

Then, consider the case when $\alpha < \alpha_0(\sigma)$ and $\sigma > \sigma_0$. In this case, we know $\beta_* < \beta^* < 1$ by Lemma 4. Further, by the same lemma and their definitions, we know that $u_{l\beta^*} \geq 0$, $v_{l\beta^*} > v_0$, and $\pi_{l\beta^*} \geq \pi_0$. Therefore, for the β^* -lenient action, the participation constraints of all the players are satisfied. Let us enforce the β^* -lenient action through coordination on z_t . There are three incentive constraints:

$$(1 - \delta)[-c + \Pr(s = 1)b + \Pr(\omega = D, s = 0)\beta^*b] + \delta u_{l\beta^*} \geq 0, \quad (IC_{\mathcal{A}}^{\beta^*})$$

$$(1 - \delta)\Pr(V_{\mathcal{P},1} = Y \mid r = 1, \beta^*)Y + \delta \pi_{l\beta^*} \geq \pi_0, \text{ and} \quad (IC_{\mathcal{P}}^{\beta^*})$$

$$\begin{aligned} (1 - \delta)v_0 + \delta v_{l\beta^*} &\geq (1 - \delta)y + \delta [\Pr(V_{\mathcal{P}} = Y \mid s = 0)v_{l\beta^*} + \Pr(V_{\mathcal{P}} = 0 \mid s = 0)v_0] \\ &\geq (1 - \delta)y + \delta \left[\frac{1 - \sigma}{2 - \sigma} v_{l\beta^*} + \frac{1}{2 - \sigma} v_0 \right]. \end{aligned} \quad (IC_{\mathcal{M}}^{\beta^*})$$

$(IC_{\mathcal{A}}^{\beta^*})$ indicates that the agent is willing to exert effort. $(IC_{\mathcal{P}}^{\beta^*})$ indicates that the principal is willing to follow the manager's recommendation (note that the principal cannot observe z). Finally, $(IC_{\mathcal{M}}^{\beta^*})$ indicates that upon $z > \beta_*$, the manager is willing to send $r = 0$ when $(\omega, s) = (D, 0)$.

Observe that $(IC_{\mathcal{A}}^{\beta^*})$ is equivalent to $u_{l\beta^*} \geq 0$. Hence, it is satisfied. From the decomposition

$$\pi_{l\beta^*} = \Pr(r = 1)\Pr(V_{\mathcal{P}} = Y \mid r = 1, \beta^*)Y + \Pr(r = 0)\pi_0$$

and $\pi_{l\beta^*} \geq \pi_0$ we know that $\Pr(V_{\mathcal{P}} = Y \mid r = 1, \beta^*)Y \geq \pi_0$. Then, as $\pi_{l\beta^*} \geq \pi_0$, $(IC_{\mathcal{P}}^{\beta^*})$ is satisfied. For the manager's incentive constraint, observe that the difference between the left-hand side and the right-hand side is continuous and increasing in δ , which is strictly positive when $\delta = 1$ (recall that $v_{l\beta^*} > v_0$) and strictly negative when $\delta = 0$. Hence, there exists $\underline{\delta}_1 \in (0, 1)$ such that $(IC_{\mathcal{M}}^{\beta^*})$ holds if and only if $\delta \geq \underline{\delta}_1$. Since $v_{l\beta^*} > v_{l\beta_*}$, comparing $(IC_{\mathcal{M}}^{\beta^*})$ with $(IC_{\mathcal{M}}^{\beta_*})$ shows $\underline{\delta}_1 < \bar{\delta}$. This completes our proof for the case when $\alpha < \alpha_0(\sigma)$. $\square$

The next lemma shows that if $\sigma > \sigma_0$, $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ may be sustained as an equilibrium payoff by non-stationary contracts, even when the players are not sufficiently patient ($\delta < \bar{\delta}$).

Lemma 8. Suppose $\alpha < \alpha_1(\sigma)$ and $\sigma > \sigma_0$. There exists $\underline{\delta} \in [\underline{\delta}_1, \bar{\delta})$ (which depends on α and σ) such that if $\delta \in [\underline{\delta}, \bar{\delta})$, $\pi_{l\beta_*}$ can be sustained by the following relational contract:

- (i) It starts with a β -lenient action in the first period for some fixed $\beta \in [0, \beta_*)$.
- (ii) It randomizes between the β -lenient action and the β^* -lenient action in the subsequent periods.
- (iii) Once the β^* -lenient action is realized, the future actions are always β^* -lenient.

Proof. We intend to find $\underline{\delta} \in [\underline{\delta}_1, \bar{\delta})$ such that if $\delta \in [\underline{\delta}, \bar{\delta})$, the equilibrium can be constructed recursively as follows. Use the profile of continuation payoffs as the state variable. Initially, the state is

$(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$, as this is the equilibrium payoff. In this state, the players enforce some β -lenient action, where $\beta \in [0, \beta_*)$. The equilibrium payoffs then decompose into the flow payoffs $(u_{l\beta}, v_{l\beta}, \pi_{l\beta})$ from the β -lenient action and the continuation payoffs (u_c, v_c, π_c) :

$$\begin{cases} u_{l\beta_*} = (1 - \delta)u_{l\beta} + \delta u_c \\ v_{l\beta_*} = (1 - \delta)v_{l\beta} + \delta v_c \\ \pi_{l\beta_*} = (1 - \delta)\pi_{l\beta} + \delta \pi_c, \end{cases}$$

where (u_c, v_c, π_c) corresponds to a β' -lenient action for some fixed $\beta' \in (\beta_*, \beta^*)$. In the next period, the state randomizes between $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ and $(u_{l\beta^*}, v_{l\beta^*}, \pi_{l\beta^*})$: it equals $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ with probability $\frac{\beta^* - \beta'}{\beta^* - \beta_*}$ and $(u_{l\beta^*}, v_{l\beta^*}, \pi_{l\beta^*})$ with probability $\frac{\beta' - \beta_*}{\beta^* - \beta_*}$. When $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ is realized, the players repeat their behavior in the initial state. When $(u_{l\beta^*}, v_{l\beta^*}, \pi_{l\beta^*})$ is realized, the players choose the β^* -lenient action, and the state stays there forever. **Figure 7** illustrates our construction.

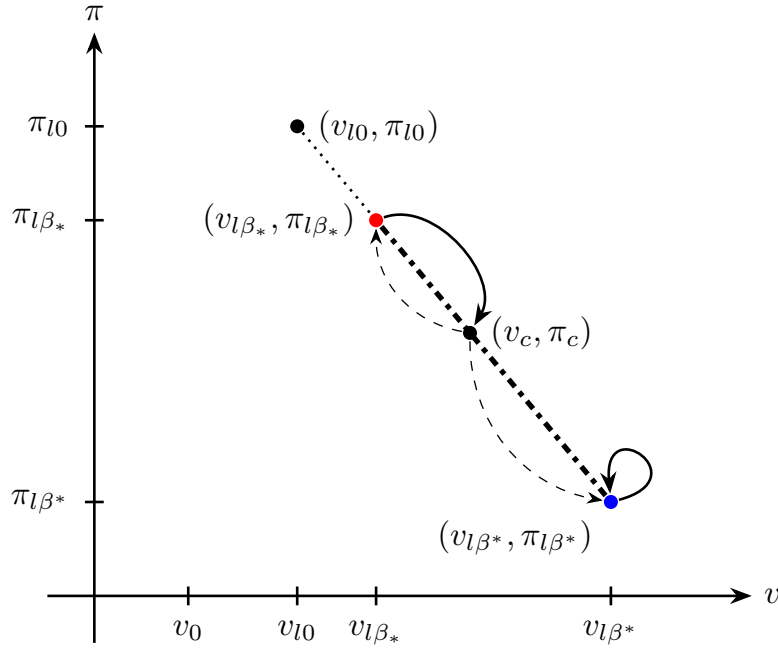


Figure 7: A non-stationary contract sustaining $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ when $\alpha < \alpha_1(\sigma)$ and $\sigma > \sigma_0$

Suppose $\sigma > \sigma_0$ and $\alpha < \alpha_1(\sigma)$. According to **Lemma 7**, the β^* -lenient action is self-enforcing if $\delta \geq \underline{\delta}_1$. Therefore, our construction reduces to the analysis of the initial state. By **Lemma 4**, we have $u_{l\beta_*} \geq 0$ and $v_{l\beta_*} > v_0$ for $\sigma > \sigma_0$. Further, as $\beta_* < \beta^*$ for $\sigma > \sigma_0$, $\pi_{l\beta_*} > \pi_{l\beta^*} \geq \pi_0$. Hence, all the participation constraints are satisfied in the initial state. It remains to show the β -lenient action and how to enforce this action.

First, in addition to $\sigma > \sigma_0$ and $\alpha < \alpha_1(\sigma)$, suppose $\alpha \geq \alpha_0(\sigma)$. In this case, let us enforce the β -lenient action through coordination on x_1 . We intend to show that there exist $\beta \in [0, \beta_*)$ and

$\underline{\delta} \in (0, \bar{\delta})$ such that when $\delta \in [\underline{\delta}, \bar{\delta})$, (u_c, v_c, π_c) is between $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ and $(u_{l\beta^*}, v_{l\beta^*}, \pi_{l\beta^*})$, and the following incentive constraints are satisfied:

$$(1 - \delta)[-c + \Pr(s_1 = 1)b + \Pr(\omega_1 = D, s_1 = 0)\beta b] + \delta u_c \geq 0, \quad (IC_{\mathcal{A}}^\beta)$$

$$(1 - \delta) \Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, x_1 \leq \beta)Y + \delta \pi_c \geq \pi_0, \quad (IC_{\mathcal{P},1}^\beta)$$

$$(1 - \delta) \Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, x_1 > \beta)Y + \delta \pi_c \geq \pi_0, \text{ and} \quad (IC_{\mathcal{P},2}^\beta)$$

$$\begin{aligned} (1 - \delta)v_0 + \delta v_c &\geq (1 - \delta)y + \delta [\Pr(V_{\mathcal{P},1} = Y \mid s_1 = 0)v_c + \Pr(V_{\mathcal{P},1} = 0 \mid s_1 = 0)v_0] \\ &\geq (1 - \delta)y + \delta \left[\frac{1 - \sigma}{2 - \sigma} v_c + \frac{1}{2 - \sigma} v_0 \right]. \end{aligned} \quad (IC_{\mathcal{M}}^\beta)$$

$(IC_{\mathcal{A}}^\beta)$ indicates that the agent is willing to exert effort. $(IC_{\mathcal{P},1}^\beta)$ and $(IC_{\mathcal{P},2}^\beta)$ indicate that the principal is willing to follow the manager's recommendation. Finally, $(IC_{\mathcal{M}}^\beta)$ indicates that upon $x_1 > \beta_*$, the manager is willing to send $r_1 = 0$ when $(\omega_1, s_1) = (D, 0)$.

To pin down β , write (u_c, v_c, π_c) as

$$u_c = \frac{u_{l\beta_*} - (1 - \delta)u_{l\beta}}{\delta}, v_c = \frac{v_{l\beta_*} - (1 - \delta)v_{l\beta}}{\delta}, \text{ and } \pi_c = \frac{\pi_{l\beta_*} - (1 - \delta)\pi_{l\beta}}{\delta}.$$

Notice that $u_{l\beta}$ and $v_{l\beta}$ are strictly increasing in β , and $\pi_{l\beta}$ is strictly decreasing in β . Hence, as β converges to β_* from below, u_c converges to $u_{l\beta_*}$ from above, v_c converges to $v_{l\beta_*}$ from above, and π_c converges to $\pi_{l\beta_*}$ from below. Then, given $\delta = \bar{\delta}$, there exists $\beta^* < \beta < \beta_*$ such that $u_c \in (u_{l\beta_*}, u_{l\beta^*})$, $v_c \in (v_{l\beta_*}, v_{l\beta^*})$, and $\pi_c \in (\pi_{l\beta^*}, \pi_{l\beta_*})$. Fix any such β . As (u_c, v_c, π_c) is continuous in δ , there exists $\underline{\delta}_2 < \bar{\delta}$ such that when $\delta \in [\underline{\delta}_2, \bar{\delta})$, $u_c \in (u_{l\beta_*}, u_{l\beta^*})$, $v_c \in (v_{l\beta_*}, v_{l\beta^*})$, and $\pi_c \in (\pi_{l\beta^*}, \pi_{l\beta_*})$. That is, for the fixed β , (u_c, v_c, π_c) is indeed between $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ and $(u_{l\beta^*}, v_{l\beta^*}, \pi_{l\beta^*})$ if $\delta \in [\underline{\delta}_2, \bar{\delta})$.

Then, we look into the incentive constraints. Observe that $(IC_{\mathcal{A}}^\beta)$ is equivalent to

$$(1 - \delta)u_{l\beta} + \delta u_c \geq 0.$$

Since $(1 - \delta)u_{l\beta} + \delta u_c = u_{l\beta_*} \geq 0$, $(IC_{\mathcal{A}}^\beta)$ is satisfied. For the principal's constraints, from the decomposition

$$\pi_{l\beta} = \Pr(r_1 = 1, x_1 \leq \beta) \Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, x_1 \leq \beta)Y + \Pr(r_1 = 0, x_1 \leq \beta)\pi_0$$

and $\pi_{l\beta} > \pi_{l\beta_*} > \pi_{l\beta^*} \geq \pi_0$ we know that $\Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, x_1 \leq \beta)Y \geq \pi_0$. Then, because $\pi_c > \pi_{l\beta^*} \geq \pi_0$, $(IC_{\mathcal{P},1}^\beta)$ is satisfied. And since $\Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, x_1 \leq \beta)Y \leq \Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, x_1 > \beta)Y$, $(IC_{\mathcal{P},2}^\beta)$ is also satisfied. For the manager's incentive constraint, observe that the difference between the left-hand side and the right-hand side is continuous and increasing in δ , which is strictly positive when $\delta = 1$ ($v_c > v_{l\beta_*} > v_0$) and strictly negative when $\delta = 0$. Hence, there exists $\underline{\delta}_3 \in (0, 1)$ such that $(IC_{\mathcal{M}}^\beta)$ holds if and only if $\delta \geq \underline{\delta}_3$. As $v_c > v_{l\beta_*}$ and $(IC_{\mathcal{M}}^\beta)$ is

easier to satisfy when v_c is larger, comparing this incentive constraint with $(IC_{\mathcal{M}}^{\beta_*})$ shows $\underline{\delta}_3 < \bar{\delta}$.

Now, instead of assuming $\alpha \geq \alpha_0(\sigma)$, consider the case when $\alpha < \alpha_0(\sigma)$. In this case, we follow the same construction as above to define β , $\underline{\delta}_2$, and $\underline{\delta}_3$. However, since $\Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, x_1 \leq \beta_*)Y < \pi_0$, $(IC_{\mathcal{P},1}^{\beta_*})$ may be violated.

We can resolve this issue with the private randomizing device z_t . Instead of x_t , let us enforce the β -lenient action through coordination on z_t . Then, both $(IC_{\mathcal{A}}^{\beta_*})$ and $(IC_{\mathcal{M}}^{\beta_*})$ remain the same. However, as the principal cannot observe z_1 , $(IC_{\mathcal{P},1}^{\beta_*})$ becomes

$$(1 - \delta) \Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, \beta_*)Y + \delta \pi_c \geq \pi_0.$$

Since $\pi_{l\beta_*} \geq \pi_0$ and $\pi_{l\beta_*} = \Pr(r_1 = 1) \Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, \beta_*)Y + \Pr(r_1 = 0) \pi_0$, we have $\Pr(V_{\mathcal{P},1} = Y \mid r_1 = 1, \beta_*)Y \geq \pi_0$. Together with $\pi_c \geq \pi_0$, this implies that the above incentive constraint is satisfied. With this fact established, the argument in the previous construction follows.

Finally, let $\underline{\delta} := \max\{\underline{\delta}_1, \underline{\delta}_2, \underline{\delta}_3\}$. Note that $\underline{\delta} < \bar{\delta}$. For arbitrarily chosen $\delta \in [\underline{\delta}, \bar{\delta})$, denote as β' the lenience level that induces $v_c = v_{l\beta'}$. As

$$\beta' = \frac{\beta^* - \beta'}{\beta^* - \beta_*} \beta_* + \frac{\beta' - \beta_*}{\beta^* - \beta_*} \beta^*,$$

we can verify that the proposed randomization between $(u_{l\beta_*}, v_{l\beta_*}, \pi_{l\beta_*})$ and $(u_{l\beta^*}, v_{l\beta^*}, \pi_{l\beta^*})$ indeed generates (u_c, v_c, π_c) . This completes our construction. $\square$

Now we can prove **Proposition 4**.

Proof of Proposition 4. Suppose $\alpha < \alpha_1(\sigma)$ and $\sigma \geq \sigma_0$. Part (i) is immediate from **Lemma 5** and **Lemma 6**. For part (ii), the result is immediate from **Lemma 5**, **Lemma 6**, and **Lemma 8**. $\square$