

Governing Informal Incentives: AI Capability and Performance Measure Design in Hierarchical Organizations *

YAO HUANG

JIN LI

LINGFANG (IVY) LI

September 22, 2026

Abstract

AI expands firms' capacity to measure performance, but these measures operate through middle managers who sustain relational contracts with workers. We study how these relationships shape performance measure design. A measure better aligned with headquarters can lower the manager's payoff from formal incentives alone, making harmful relational contracts easier to sustain. Headquarters may therefore deliberately limit the use of better-aligned performance measures. Improvements in AI-enabled measurement can thus yield limited gains even under optimal design.

Keywords: Performance Measurement, Hierarchies, Relational Contracts, Artificial Intelligence

JEL code: D23, D86, L22, M52

*Yao Huang: The University of Hong Kong. E-mail: yaohuang@connect.hku.hk. Jin Li: The University of Hong Kong. E-mail: jli1@hku.hk. Lingfang (Ivy) Li: Fudan University. E-mail: lingfangli@fudan.edu.cn. We especially thank Bob Gibbons for conversations and detailed comments. We thank Daniel Barron, Alessandro Bonatti, Bonnie Cao, Limei Chen, Yifan Dai, Wouter Dessein, Matthias Fahn, Jie Gong, Xinyu Hua, Peixuan Huang, Fang Li, Zhen Liang, Kori Liu, Zhuoran Lu, Yuchen Luo, Zhiwei Shang, Yanhui Wu, Xuefei Yu, Yiyang Zhou, and seminar participants at Fudan, HKU, LEOH, and MIT for helpful comments and suggestions. Lingfang (Ivy) Li acknowledges financial support from the National Natural Science Foundation of China (72192845, 72131004). The authors are listed in alphabetical order and are all first authors.

1 Introduction

Advances in artificial intelligence (AI) create new opportunities to improve firm performance, but realizing these gains depends on how firms organize work. Survey evidence identifies organizational and cultural factors as leading obstacles to AI adoption (Li, Zhu and Hua, 2025). One key determinant of these factors is middle managers, who play a central role in implementing AI (Tsipursky, 2026). These managers translate organizational objectives into daily work and build relationships that motivate employees (Field, Hancock and Schaninger, 2023). They exercise discretion over how objectives apply to their units and what rewards to promise workers. That discretion also allows them to pursue goals that differ from headquarters’.

We study how this discretion shapes the gains from using AI to improve performance measurement. AI allows firms to track performance in real time and capture dimensions of worker behavior that were once difficult to observe (Leonardi and Contractor, 2018; de la Garza, 2019; Kellogg, Valentine and Christin, 2020). The effects these new measures create, however, depend on how middle managers incorporate them into existing systems of evaluation and rewards.

Dingdong, an online grocery retailer, illustrates this situation. Its AI-assisted Voice of Customer (VOC) system aggregates customer feedback, classifies complaints, attributes them to units or workers, and feeds the resulting measures into KPIs. Headquarters defines the overall metrics, while department managers translate them into standards and weights for their units. VOC is also only one input into reviews and promotions: managers retain discretion to recognize contributions that the scores miss (Li, Chen and Liu, 2025; Li et al., 2026). How, then, do middle managers shape the gains from better measurement, and how should headquarters design measures in response?

We study a hierarchy with headquarters, a manager, and a worker who allocates effort across activities. Headquarters and the manager prefer different effort allocations. The manager is essential in the baseline, so headquarters cannot contract with the worker directly. Headquarters chooses a formal performance measure from a set determined by the firm’s AI capability. The manager sets a bonus tied to this measure and may also promise a discretionary bonus for contributions to her own goal. The manager and the worker interact repeatedly, allowing them to sustain a relational contract.

The manager uses relational contracting to pull the worker’s effort away from the measured direction and toward her own goal. Whether this serves headquarters depends on which is closer to headquarters’ goal, the formal performance measure or the manager’s goal. The relational

contract helps headquarters when the formal measure is the less aligned of the two and hurts headquarters when the measure is the more aligned.

In either case, the manager's ability to sustain a relational contract depends on her fallback position: what she can earn using formal incentives alone after reneging on a promised bonus. A lower fallback increases what she loses from a breakdown in cooperation and allows her to sustain a stronger relational contract. The formal measure headquarters designs therefore determines the strength of the informal incentives the manager can sustain.

At low AI capability, the manager's goal is closer to headquarters' goal than any feasible formal performance measure. The relational contract therefore benefits headquarters. Better formal measurement can raise the manager's fallback and crowd out this beneficial relational contract, following the familiar logic of [Baker, Gibbons and Murphy \(1994\)](#). Headquarters optimally withholds formal measurement to preserve the beneficial relationship.

Our main result arises at high capability, when better measurement can instead enable a harmful relational contract. Once the formal measure is better aligned with headquarters' goal than the manager's goal is, further alignment makes formal contracting less attractive to the manager. Her fallback deteriorates. If the fallback becomes sufficiently unattractive, she can credibly promise a bonus that redirects effort toward her own goal. A formal measure that better reflects headquarters' goal can thus reduce headquarters' payoff by enabling the harmful relationship.

Anticipating this, headquarters responds by choosing a deliberately imperfect performance measure. It stops short of the most aligned feasible formal measure to keep the manager's fallback attractive enough to prevent the harmful relational contract. Relational contracting, therefore, limits the returns to better performance measurement.

When headquarters designs the performance measure optimally, its payoff is flat at low and high levels of AI capability and rises only in between, where headquarters adopts the most aligned feasible measure without enabling a relational contract. The two flat regions have different sources: preserving a beneficial relationship at low capability and preventing a harmful one at high capability. Within a fixed hierarchy, technological progress can expand what is measurable without improving what headquarters can achieve.

Further gains then require complementary organizational change. The Supplemental Appendix illustrates one possibility: headquarters may remove the managerial layer as AI capability improves. The analysis thus connects to organizational complementarities ([Brynjolfsson and Milgrom, 2013](#)) and provides a mechanism through which changes in organizational design can

be necessary to realize the gains from AI (Brynjolfsson, Rock and Syverson, 2021).

Our paper connects to the literature on performance measurement and relational contracts. Imperfect formal measures can distort effort allocation (Kerr, 1975; Holmström and Milgrom, 1991), and relational contracts can mitigate these distortions, with their credibility depending on the parties' formal fallback (Baker, Gibbons and Murphy, 1994, 1999, 2002, 2023). In these two-party settings, optimal formal design often supports valuable relational contracts by reducing the temptation to renege. We add a hierarchy in which the party controlling the relational contract is not the party designing the measure. The new implication is that headquarters may refrain from using better measures to prevent detrimental relational contracts between the manager and the worker.

Our paper also relates to the literature on authority and incentives in hierarchies.¹ A principal may hold formal authority while a better-informed subordinate exercises real control (Aghion and Tirole, 1997). A supervisor who is better informed than the principal may collude with the agent, and the principal can respond with collusion-proof contracts (Tirole, 1986), auditors (Kofman and Lawarrée, 1993), or the choice of organizational form (Faure-Grimaud, Laffont and Martimort, 2003). In these models the subordinate's leverage comes from private information. In ours it comes from the relational contract with the worker.

Recent work incorporates relational contracts into hierarchies.² Closest to our paper is Troya-Martinez and Wren-Lewis (2023), who also study a principal who governs manager-agent relational contracts. They show that the relational contracts can help the principal by strengthening the manager's commitment, but can also support collusion. The principal governs this relationship through the manager's profit share and a cap on the price the manager can pay the agent. We focus on a different governance instrument: headquarters designs the performance measure the manager uses with the worker, which shapes not only formal incentives but also the manager's fallback after reneging, and thus the relational contract she can sustain.

Lastly, work on AI in organizations studies AI's effects on tasks (Acemoglu and Restrepo, 2019), productivity (Brynjolfsson, Li and Raymond, 2025), worker motivation (Tong et al., 2021; Chen et al., 2025), job design (Fahn, Li and Sun, 2026), decision authority (Athey, Bryan and Gans, 2020), and the structure of hierarchies (Ide and Talamàs, 2025). We study AI as a measurement technology. By expanding what a firm can measure, AI changes not only worker incentives but also the manager's informal control. That change can limit the value of better measurement.

¹See Bolton and Dewatripont (2013) and Mookherjee (2013) for reviews.

²See, e.g., Fong and Li (2017), Choe and Ishiguro (2022), and Li et al. (2025).

2 Model Setup

We develop a model of a three-tier organization composed of headquarters, a manager, and a worker. The model has two stages. In the first stage, headquarters designs a formal performance measure. In the second stage, the manager uses the measure to motivate the worker within a long-term relationship. We model the relationship between the manager and the worker as an infinitely repeated game à la [Baker, Gibbons and Murphy \(1994\)](#).

For clarity, we present these two stages in reverse order. We begin by describing the second stage: the interaction between the manager and the worker. We first describe their single-period (stage) game, then extend this to their repeated game and the strategies they employ. After establishing this, we return to the first stage of the model, where headquarters designs the performance measure that will govern the manager-worker interaction.

The Stage Game In describing the stage game, we omit time subscripts for convenience.

Effort and Value Creation. In each period the worker chooses an unobservable, two-dimensional effort $e = (e_1, e_2) \geq 0$ at a cost $c(e) = \frac{1}{2}(e_1^2 + e_2^2)$. This effort stochastically determines the worker's contribution to headquarters' value y_h and the manager's value y_m . Assume that y_h and y_m are binary, taking a value of either 1 (high) or 0 (low).

Headquarters and the manager value the worker's effort differently. Each party has a preferred effort direction, which we call its *goal* and represent by an angle: θ_h for headquarters and θ_m for the manager. When the worker's effort aligns more closely with party i 's goal, a high value for that party is more likely:

$$\Pr[y_i = 1|e] = \cos(\theta_i) e_1 + \sin(\theta_i) e_2, \quad i \in \{h, m\}.$$

We normalize $\theta_h = 0$ and, for expositional simplicity, set $\theta_m = \frac{\pi}{4}$. Headquarters therefore values only the first dimension of effort, while the manager values both dimensions equally. Figure 1 shows these two goals in the (e_1, e_2) plane.

Performance Measurement and Compensation. As in [Baker, Gibbons and Murphy \(1994\)](#), we assume that the true value of headquarters, y_h , is observable but not contractible. Instead, formal contracts must depend on an imperfect performance measure designed by headquarters at the outset.

We represent the performance measure by an angle $\theta_p \in [0, \frac{\pi}{2}]$. In each period, this measure yields an objective performance outcome p , which takes 1 (high) or 0 (low). The probability of a

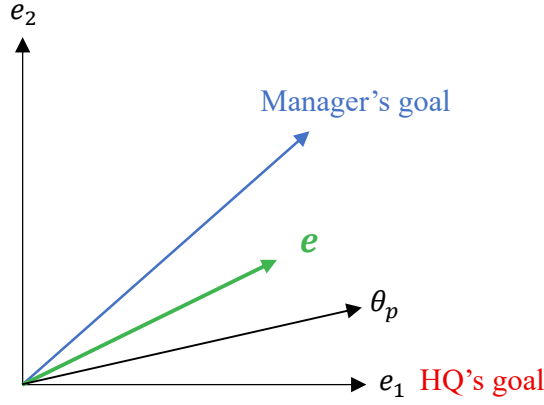


Figure 1: Goals for Headquarters and the Manager, and the Performance Measure.

high performance outcome

$$\Pr[p = 1|e] = \cos(\theta_p)e_1 + \sin(\theta_p)e_2.$$

Therefore, the performance outcome (p) is more likely to be high when the worker's effort (e) aligns more closely with the headquarters-designed performance measure (θ_p).

To motivate the worker, the manager offers a compensation package. This package includes a fixed salary s , a bonus b contingent on the worker's contribution to the manager's value (y_m), and a bonus β contingent on the performance outcome (p). The worker's total compensation is then $w = s + by_m + \beta p$.

The two bonuses differ in their enforceability. Because the performance outcome (p) can be objectively measured, the bonus β is a formal contract enforceable by a court. In contrast, we assume that the manager's value (y_m) cannot be objectively measured, and therefore, cannot be verified by the court. Consequently, the bonus b is a relational contract, paid at the discretion of the manager.

Payoffs. The per-period payoffs for the three players are as follows. Headquarters' payoff is its value, y_h . The manager's payoff is $y_m - w$, which is the manager's value less the worker's compensation. The worker's payoff is $w - c(e)$, which is the worker's compensation less the cost of effort.

Timing of the Stage Game. The sequence of events within each period is as follows. First, the manager offers the worker a compensation package (s, b, β) . Second, the worker either accepts the contract or rejects it. If the worker rejects, both the worker and the manager receive zero in that period. Third, if the worker accepts, then the worker chooses an effort e . Fourth, the manager and the worker observe the realization of the worker's contribution (y_m), and the manager and the

worker (and, if necessary, a court) observe the realization of the objective performance outcome (p). Finally, if $p = 1$ the manager pays the bonus β dictated by the formal contract, and if $y_m = 1$ the manager chooses whether to pay the worker the bonus b indicated by the relational contract.

The Repeated Game and Strategies The manager and the worker play the stage game repeatedly with a discount rate r . In the repeated game, we follow [Baker, Gibbons and Murphy \(1994\)](#) and consider stationary equilibria in which the manager and the worker play the following trigger strategies. The manager and the worker begin by cooperating: the worker trusts the manager's promise to pay the relational bonus b , and the manager honors this promise. They continue to cooperate as long as the manager never reneges. If the manager ever reneges, they permanently revert to the equilibrium formal contract in the one-shot game. This fallback position yields the manager an expected per-period payoff of v^{spot} , which we will derive later.

Headquarters' Problem We now formally state the optimization problem for headquarters and the manager. We begin with the manager's problem for a given performance measure and then turn to headquarters' problem of designing the performance measure.

The Manager's Problem. For a given performance measure θ_p , the manager chooses the compensation package (s, b, β) to maximize the manager's expected payoff:

$$\max_{s, b, \beta} v = \mathbb{E}[y_m - w | e(b, \beta)],$$

subject to the following constraints.

(i) Incentive Compatibility: The worker must be incentivized to provide effort. The worker's chosen effort will be the one that maximizes the worker's own expected payoff:

$$e(b, \beta) = \operatorname{argmax}_e \mathbb{E}[w | e] - c(e),$$

(ii) Participation: The worker must be willing to accept the contract. The worker's expected payoff must be greater than zero:

$$\mathbb{E}[w | e(b, \beta)] - c(e(b, \beta)) \geq 0,$$

(iii) Non-reneging: For the relational contract to be feasible, the manager must have an incentive to honor the promise. The one-time gain from reneging on the promised bonus must be less than the present value of the future payoffs lost from destroying the relationship:

$$b \leq \frac{1}{r}(v - v^{\text{spot}}).$$

Headquarters' Problem. In the first stage, headquarters designs the performance measure θ_p to maximize its expected payoff. Headquarters does so with anticipation of how this choice will shape the optimal compensation package the manager offers to the worker. Formally, headquarters chooses θ_p to maximize its expected payoff:

$$\max_{\theta_p} \mathbb{E}[y_h] = e_1,$$

subject to the constraint that e_1 is the first dimension of the effort vector e resulting from the manager's optimization problem defined above. To make the relational contract between the manager and worker relevant in the performance measure design, we assume that the discount rate r is smaller than a threshold $\bar{r}$.³

Remarks We conclude this section with three remarks on the model.

First, we summarize the measurement system by a single incentive direction, θ_p . This does not mean that the system uses only one measure. Dingdong's measurement system, for example, combines multiple performance measures. The angle θ_p captures the effort direction rewarded by the resulting formal system. Allowing multiple measures does not change the main results.

Second, we assume headquarters maximizes its own value y_h . This objective is consistent with the interpretation that headquarters allocates a fixed budget to the manager's division, and the manager pays the worker out of this budget. Alternatively, headquarters could maximize total surplus. The Supplemental Appendix shows that our results remain qualitatively unchanged.

Finally, we assume that headquarters designs the performance measure while the manager pays the bonus. Alternative settings could shift this balance of power: the manager could design the performance measure (giving the manager more authority) or headquarters could pay the bonus (giving headquarters more control). In practice, the relationship between headquarters and managers varies considerably across organizations (see [Chandler, 1977](#)). Our model takes a middle path between these extremes.

3 Preliminary Analysis: The Manager-Worker Relationship

In this section, we characterize the optimal formal and relational contracts that the manager offers to the worker for a given performance measure θ_p . We show that the manager's choice of contracts

³If $r > \bar{r}$, the relational contract is not a binding constraint in the performance measure design, and the problem simplifies to a measure design problem where the manager can only use a formal contract. We provide the expression of $\bar{r}$ in the Appendix.

depends on the distortion in the performance measure from the manager's perspective.

For the manager, the performance measure θ_p is distorted whenever θ_p deviates from the manager's goal ($\frac{\pi}{4}$). We define the *distortion in the performance measure* by $\theta = |\theta_p - \frac{\pi}{4}|$. The following lemma shows how this distortion determines the optimal formal and relational contracts.

Lemma 1. *The optimal formal (β) and relational (b) bonuses are functions of the distortion (θ):*

- (i) *If $\theta \leq \hat{\theta}$, the manager relies solely on the formal contract ($b = 0$), with β decreasing in θ .*
- (ii) *If $\theta > \hat{\theta}$, the manager also uses the relational contract ($b > 0$), with b increasing and β decreasing in θ .*

The threshold $\hat{\theta}$ increases in the discount rate r .

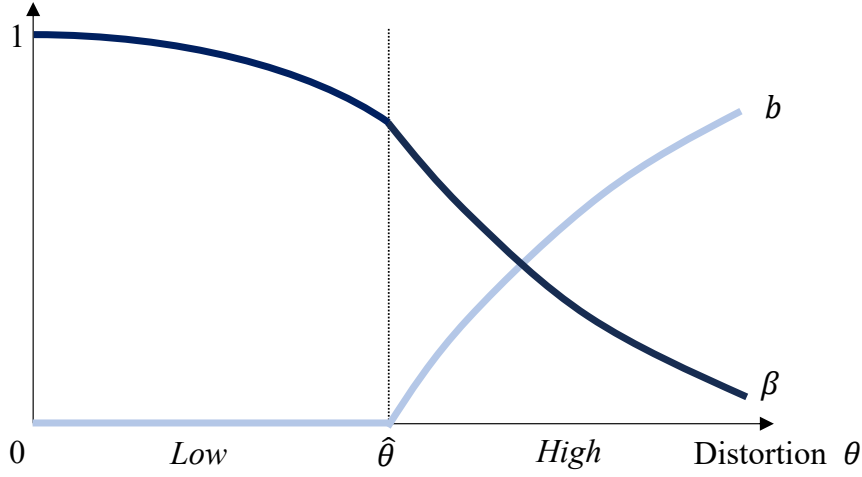
Lemma 1 shows that whether the manager can sustain a relational contract depends on the distortion in the performance measure (Figure 2a). When distortion is high ($\theta > \hat{\theta}$), the manager can sustain a relational contract ($b > 0$). When distortion is low ($\theta \leq \hat{\theta}$), the manager cannot sustain any relational contract ($b = 0$) and must rely solely on the formal contract.

As in Baker, Gibbons and Murphy (1994), the feasibility of the relational contract depends on the attractiveness of the manager's fallback position after reneging. Recall that the fallback is v^{spot} that the manager earns by reverting to the formal contract alone. The formal contract is more effective for the manager when the performance measure is less distorted. Consequently, the fallback payoff v^{spot} is decreasing in θ .⁴

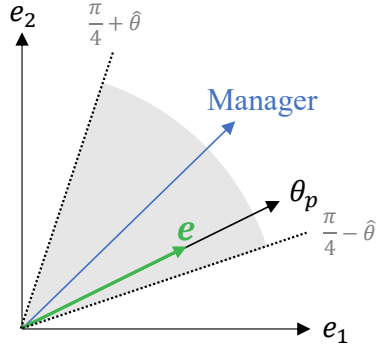
Given the payoffs from honoring (v) and from reneging on (v^{spot}) the discretionary bonus b , the manager honors the relational contract if and only if the present value of the relationship exceeds the size of the bonus: $(v - v^{\text{spot}})/r \geq b$. When distortion is high ($\theta > \hat{\theta}$), the fallback payoff v^{spot} is low. The present surplus of the ongoing relationship, $(v - v^{\text{spot}})/r$, is then large, so the manager can credibly sustain a relational bonus. If the distortion decreases, v^{spot} improves and reduces the present value of the relationship. The manager can then sustain only a smaller relational bonus and relies more heavily on the formal contract. Once distortion becomes sufficiently low ($\theta \leq \hat{\theta}$), the fallback position is so attractive that the manager will renege on any relational contract ($b = 0$).

The worker's effort allocation, therefore, depends on the performance measure through its effect on the optimal contracts. When the distortion in the performance measure is low ($\theta \leq \hat{\theta}$),

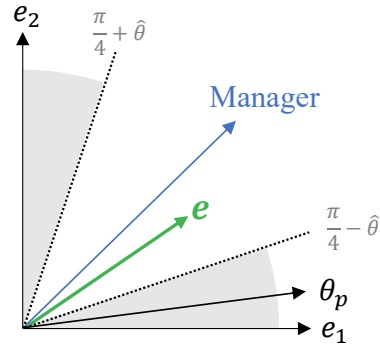
⁴Formally, the equilibrium formal contract in the one-shot game yields the manager an expected per-period payoff of $v^{\text{spot}}(\theta) = \frac{1}{2} \cos^2(\theta)$, which increases as θ decreases.



(a) Optimal formal (β) and relational (b) bonuses



(b) Effort under low distortion ($\theta \leq \hat{\theta}$)



(c) Effort under high distortion ($\theta > \hat{\theta}$)

Figure 2: Distortion Determines the Optimal Bonuses and the Worker's Effort

the manager offers the formal contract only. Consequently, the worker's effort aligns with the formal performance measure, θ_p (Figure 2b). In contrast, when the distortion is high ($\theta > \hat{\theta}$), the manager also relies on the relational contract. Because the relational contract rewards the effort that aligns with the manager's goal ($\frac{\pi}{4}$), it pulls the worker's effort away from θ_p and toward $\frac{\pi}{4}$ (Figure 2c).

The preceding analysis establishes how any given performance measure shapes the manager-worker relationship and, in turn, the worker's effort. We can now address the central question of this paper: how does headquarters, anticipating this manager-worker relationship, design the performance measure?

4 AI Capability and the Design of Performance Measures

We now turn to headquarters' problem: choosing a performance measure θ_p to best align the worker's effort with headquarters' goal, located at 0. The analysis proceeds in three steps. First, we show how a performance measure affects headquarters' and the manager's payoffs. Second, we introduce the firm's AI capability, which constrains the set of feasible measures, and characterize the measure headquarters adopts. Third, we ask when greater AI capability creates value.

4.1 The Effect of Performance Measures on Headquarters and the Manager

A performance measure θ_p affects the payoffs of headquarters and the manager in two ways. Directly, the measure enters the formal contract. Indirectly, the measure shapes the relational contract through the manager's fallback. The formal contract tilts effort toward the measure, and the relational contract tilts effort toward the manager's goal.

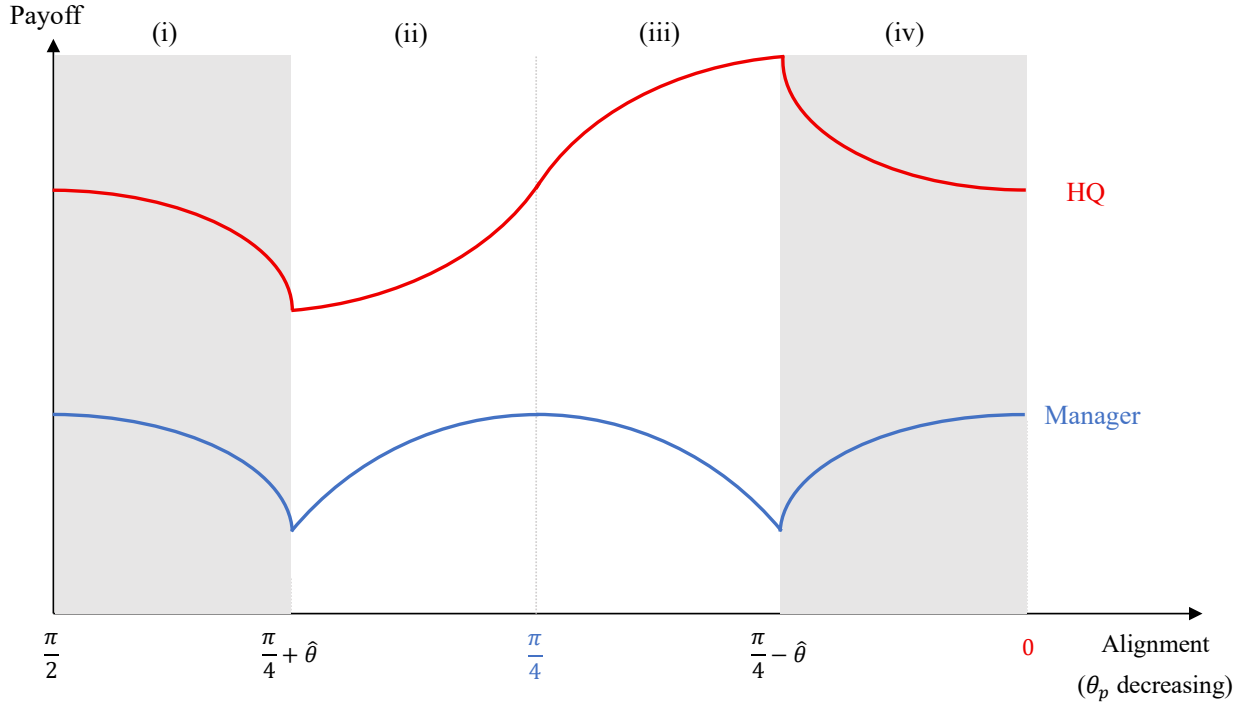
The following lemma shows how both payoffs change as the measure becomes more aligned with headquarters' goal.

Lemma 2. *As the performance measure θ_p becomes more aligned with headquarters' goal, moving from $\frac{\pi}{2}$ toward 0, the payoffs of headquarters and the manager change as follows:*

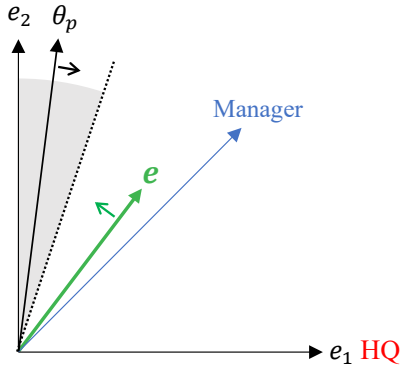
- (i) *Both payoffs decrease as θ_p decreases from $\frac{\pi}{2}$ to $\frac{\pi}{4} + \hat{\theta}$.*
- (ii) *Both payoffs increase as θ_p decreases from $\frac{\pi}{4} + \hat{\theta}$ to $\frac{\pi}{4}$.*
- (iii) *Headquarters' payoff increases and the manager's payoff decreases as θ_p decreases from $\frac{\pi}{4}$ to $\frac{\pi}{4} - \hat{\theta}$.*
- (iv) *Headquarters' payoff decreases and the manager's payoff increases as θ_p decreases from $\frac{\pi}{4} - \hat{\theta}$ to 0.*

The lemma divides the range of measures into four regions, as Figure 3a shows. In the first two regions, the interests of headquarters and the manager are aligned: as the measure improves, both payoffs fall in region (i) and both rise in region (ii). In the last two regions, their interests diverge: headquarters gains at the manager's expense in region (iii), and the manager gains at headquarters' expense in region (iv). In the figure, relational contracts emerge at both extremes, shown as the shaded regions. In between, the manager relies on the formal contract alone.

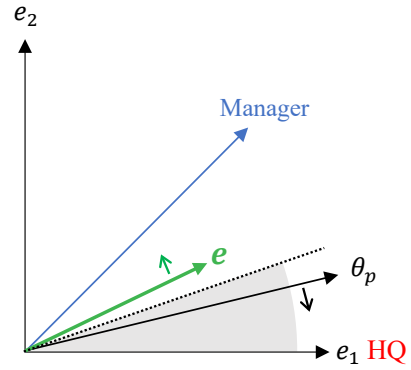
Regions (i) and (ii) follow the familiar two-party logic in [Baker, Gibbons and Murphy \(1994\)](#). The hierarchy effectively reduces to two tiers because headquarters and the manager are aligned in these regions. In region (i), the measure is far from both goals. The manager's fallback is therefore unattractive, and she sustains a relational contract that pulls effort toward her goal and



(a) Payoffs across performance measures (shaded: relational contract active)



(b) Effort in region (i)



(c) Effort in region (iv)

Figure 3: Headquarters' and the Manager's Payoffs Across Performance Measures

closer to headquarters' goal. As the measure improves, her fallback becomes more attractive and crowds out this beneficial relational contract. Effort shifts toward the formal measure, which remains distorted for both parties, so both payoffs fall (Figure 3b). In region (ii), the relational contract is gone, and further improvements move effort closer to both goals and raise both payoffs.

Regions (iii) and (iv) are where the hierarchy matters. Once the measure crosses the manager's goal, headquarters and the manager come into conflict: a measure closer to headquarters' goal is now farther from the manager's. Whether the manager can counteract through relational

contracting distinguishes the two regions.

In region (iii), the manager cannot sustain a relational contract. The measure remains close enough to her goal that her fallback is attractive, so she would renege on any relational contract. Effort therefore follows the formal measure. As the measure improves, effort shifts toward headquarters' goal and away from the manager's. Headquarters' payoff rises, while the manager's falls.

In region (iv), the manager can again sustain a relational contract, and it now works against headquarters. Once the measure moves past $\frac{\pi}{4} - \hat{\theta}$ toward headquarters' goal, it lies far enough from the manager's goal to make her fallback unattractive. The relational contract becomes credible and pulls effort away from the more aligned measure and back toward the manager's goal. As the measure improves further, her fallback becomes less attractive and the relational contract strengthens, pulling effort farther back (Figure 3c). Headquarters' payoff falls, while the manager's rises.

In short, improving formal measures first crowds out "good" relational contracts, and then, if pushed too far, creates "bad" ones. The former is familiar. The latter arises only in a hierarchy, where the manager can exploit discretion to form an informal relationship below. Headquarters must therefore design the measure not only to guide the worker, but also to govern the manager, as we next show.

4.2 Optimal Measurement Design

We now ask which measure headquarters should adopt. The set of feasible measures depends on the firm's AI capability. Define the firm's AI capability $\bar{\theta} \in [0, \frac{\pi}{2}]$ as the most aligned measure headquarters can implement. Headquarters can choose any measure $\theta_p \geq \bar{\theta}$ but cannot choose a measure below $\bar{\theta}$. A lower $\bar{\theta}$ therefore corresponds to higher capability: the firm can design measures more closely aligned with headquarters' goal.

Given the capability $\bar{\theta}$, headquarters chooses the feasible measure that maximizes its payoff. The proposition below characterizes the optimal choice.

Proposition 1. *The optimal performance measure depends on the firm's AI capability:*

- (i) *For firms with low capability ($\bar{\theta} \in (\frac{\pi}{4}, \frac{\pi}{2}]$), headquarters implements no formal measure.*
- (ii) *For firms with moderate capability ($\bar{\theta} \in [\frac{\pi}{4} - \hat{\theta}, \frac{\pi}{4}]$), headquarters implements $\theta_p^* = \bar{\theta}$.*
- (iii) *For firms with high capability ($\bar{\theta} \in [0, \frac{\pi}{4} - \hat{\theta})$), headquarters implements $\theta_p^* = \frac{\pi}{4} - \hat{\theta}$.*

The proposition shows that the optimal measure is not always the most aligned one available. For firms with low capability, headquarters withholds the measure entirely, implementing no formal measure. For firms with moderate capability, headquarters implements the most aligned available measure. For firms with high capability, headquarters restrains the measure at $\frac{\pi}{4} - \hat{\theta}$ rather than pushing it to its technological frontier.

The choice in each regime follows from Lemma 2. When capability is low, introducing any formal measure would risk crowding out a relational contract that benefits both headquarters and the manager. Headquarters therefore withholds the formal measurement and relies solely on the manager's relational contract. This aligns effort with the manager's goal $\frac{\pi}{4}$ —imperfect for headquarters, but the best headquarters can attain. The concern behind this choice echoes Fei's account of pre-1949 rural China, where the new judicial system was destroying the original ritual order but had not yet established an effective order based on law (Fei, 1992, p. 106).

When capability grows to moderate, headquarters implements the formal measure and pushes it to the technological frontier. The measure is now closer to headquarters' goal than the manager's goal is, so it directs effort more effectively than the manager could. The manager loses because effort moves away from her goal and toward headquarters' goal. But she cannot counteract the measure through relational contracting because her formal fallback remains attractive enough.

When capability becomes high, pushing the measure to the technological frontier would backfire. A measure more aligned than $\frac{\pi}{4} - \hat{\theta}$ would lie too far from the manager's goal, worsen the manager's formal fallback, and let the manager sustain a relational contract again, one that would pull effort back toward the manager's goal. Headquarters therefore restrains the measure at $\frac{\pi}{4} - \hat{\theta}$, the most aligned measure that still prevents the relational contract from forming.

We call the last strategy *strategic imperfection*: headquarters restrains its use of the technology to keep the manager from building informal control beneath the formal system. Thus, even though no relational contract forms at the optimum, the threat of one shapes the design.

The use of formal imperfection to contain informal control echoes observations from organizational studies. In *Men Who Manage*, Dalton (1959) describes the Field Work Department as a “theoretically sound” control system. Yet it was “undone by politics” because it failed to fit the informal relations among local managers and foremen. The central office's response was not to impose a more complete control system. It sought to “fit its scheme” to local managers' attitudes while shifting some control back to the office. Our result captures this logic.

The three regimes show how headquarters' stance toward the manager changes with capability. Under low capability, the manager is an ally: headquarters relies on the manager to direct the

worker. Under moderate capability, the alliance breaks and headquarters begins to implement its own formal measure to discipline the manager. Under high capability, the manager becomes a potential threat, which headquarters contains through strategic imperfection.

4.3 The Return to AI Capability

After characterizing the optimal measure for a given AI capability, we now ask how headquarters' payoff changes as capability improves. Greater capability lowers $\bar{\theta}$ and expands the set of feasible measures. The next proposition characterizes the return to AI capability.

Proposition 2. *As AI capability improves, so that $\bar{\theta}$ decreases from $\frac{\pi}{2}$ to 0, headquarters' payoff under the optimal measure θ_p^* :*

- (i) *is constant as $\bar{\theta}$ decreases from $\frac{\pi}{2}$ to $\frac{\pi}{4}$;*
- (ii) *strictly increases as $\bar{\theta}$ decreases from $\frac{\pi}{4}$ to $\frac{\pi}{4} - \hat{\theta}$;*
- (iii) *is constant as $\bar{\theta}$ decreases from $\frac{\pi}{4} - \hat{\theta}$ to 0.*

Proposition 2 shows that, as AI capability improves, headquarters' payoff stays flat at low capability, rises at moderate capability, and plateaus at high capability (Figure 4). The intuition follows from Proposition 1. When capability is low, headquarters gains nothing from improving capability: it withholds the measure and relies on the manager. Once capability crosses $\frac{\pi}{4}$, headquarters begins implementing the frontier measure ($\theta_p = \bar{\theta}$), so each further improvement raises headquarters' payoff. Yet pushing beyond $\frac{\pi}{4} - \hat{\theta}$ yields no additional return. Headquarters no longer uses the frontier measure. It holds the measure back to prevent the manager from sidestepping formal control through a relational contract.

The final plateau reflects an organizational constraint, not a technological one. Headquarters deliberately uses a less aligned measure to govern the middle layer of the hierarchy. Our results are therefore most relevant when middle managers remain necessary for reasons outside the model. Span of control may be limited, local information may matter, tasks may be hard to codify, or workers may need guidance. In such settings, headquarters cannot simply replace the middle layer with a better measure. The measure still works through the hierarchy, and the hierarchy shapes what the measure can achieve. Further gains from AI may therefore require rethinking the hierarchy itself; the Supplemental Appendix explores one option.

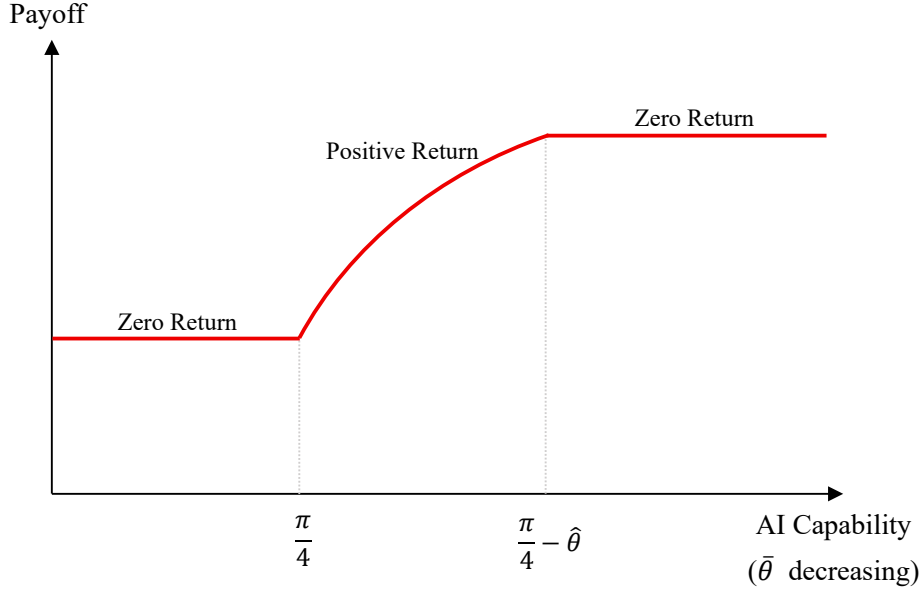


Figure 4: The Return to AI Capability

5 Conclusion

In a hierarchy, a performance measure is more than a formal incentive device. When the manager who applies it retains informal control over the worker, the measure also shapes that informal control. This paper studies how headquarters should design AI-enabled measures in such a setting. AI expands the set of formal measures headquarters can implement, but the manager sustains the relational contract with the worker. The measure therefore affects worker effort not only directly, through formal incentives, but also indirectly, through relational incentives by changing the manager's fallback payoff after reneging.

The interaction between the formal measure and the manager-worker informal relationship shapes the optimal use of AI capability. When capability is low, headquarters withholds formal measurement, because any available measure would risk crowding out a manager-worker relationship that serves headquarters. When capability is moderate, headquarters adopts the most aligned measure it can implement. When capability is high, headquarters deliberately holds back, choosing an imperfect measure tilted toward the manager. This strategic imperfection keeps the manager's fallback attractive enough to prevent a relational contract that would work against headquarters.

The return to AI capability is therefore zero when capability is low or high, and positive only in between. The zero return at high capability reflects an organizational constraint. Where formal measures must work through middle managers, realizing further value from AI may require

complementary changes in organizational design rather than better measurement alone.

Appendix

Preliminaries. Consider a compensation package (s, b, β) . If the worker believes the manager will honor the relational contract, the worker solves

$$\max_e s + b \mathbb{E}[y_m|e] + \beta \mathbb{E}[p|e] - c(e),$$

yielding

$$e(b, \beta) = b \cdot (\cos \frac{\pi}{4}, \sin \frac{\pi}{4}) + \beta \cdot (\cos \theta_p, \sin \theta_p). \quad (1)$$

The worker accepts the compensation package if the expected payoff is non-negative:

$$s + b \mathbb{E}[y_m|e(b, \beta)] + \beta \mathbb{E}[p|e(b, \beta)] - c(e(b, \beta)) \geq 0. \quad (2)$$

The manager's expected payoff per period, given b and β , is

$$\mathbb{E}[y_m|e(b, \beta)] - \{s + b \mathbb{E}[y_m|e(b, \beta)] + \beta \mathbb{E}[p|e(b, \beta)]\}. \quad (3)$$

The manager sets s at the lowest level satisfying (2). Substituting into (3) yields the manager's expected payoff $v(b, \beta)$:

$$v(b, \beta) = \mathbb{E}[y_m|e(b, \beta)] - c(e(b, \beta)) = b + \beta \cos \theta - \frac{1}{2} (b^2 + \beta^2 + 2b\beta \cos \theta),$$

where $\theta = |\theta_p - \frac{\pi}{4}|$ is the distortion.

In spot contracting, relational contracts are infeasible ($b = 0$). The manager chooses β to maximize $v(0, \beta)$, which yields $\beta^{\text{spot}} = \cos \theta$ and a fallback payoff $v^{\text{spot}} = \frac{1}{2} \cos^2 \theta$. In the repeated relationship, the manager maximizes $v(b, \beta)$ subject to the non-reneging constraint $b \leq (v(b, \beta) - v^{\text{spot}})/r$. Throughout, we assume $0 < r \leq \bar{r} := \sin^2(\frac{\pi}{8})$.

Proof of Lemma 1. Note that $v(b, \beta) = \frac{1}{2} - \frac{1}{2}(1-b)^2 \sin^2 \theta - \frac{1}{2}[\beta - (1-b) \cos \theta]^2$, so $v \leq \frac{1}{2}$.

If $\sin^2 \theta \geq 2r$, the contract $b = 1, \beta = 0$ reaches this upper bound, and it satisfies the non-reneging constraint because $v - v^{\text{spot}} = \frac{1}{2} \sin^2 \theta \geq r$.

Suppose $\sin^2 \theta < 2r$. Then $rb \leq v - v^{\text{spot}} \leq \frac{1}{2} \sin^2 \theta$ gives $b < 1$. Because β affects v only through $-\frac{1}{2}[\beta - (1-b) \cos \theta]^2$, setting $\beta = (1-b) \cos \theta$ maximizes v and hence also relaxes the non-reneging constraint, which becomes $b [(\sin^2 \theta - r) - \frac{1}{2}b \sin^2 \theta] \geq 0$. Because $v = \frac{1}{2} - \frac{1}{2}(1 -$

$b)^2 \sin^2 \theta$ increases in $b < 1$, the manager sets b at the largest value that satisfies the non-reneging constraint. This yields $b = 0$ if $\sin^2 \theta \leq r$, and $b = 2(1 - r / \sin^2 \theta)$ if $r < \sin^2 \theta < 2r$.

Consequently,

$$b(\theta) = \begin{cases} 0, & \theta \leq \hat{\theta}, \\ 2 \left(1 - \frac{r}{\sin^2 \theta} \right), & \hat{\theta} < \theta < \theta_{RC}, \text{ and } \beta(\theta) = [1 - b(\theta)] \cos \theta \\ 1, & \theta \geq \theta_{RC}, \end{cases}$$

where $\hat{\theta} := \arccos \sqrt{1 - r}$ and $\theta_{RC} := \arccos \sqrt{1 - 2r}$, both of which increase in r . Under the assumption $r \leq \sin^2(\frac{\pi}{8})$, we have $r < \frac{1}{4}$, so $0 < \hat{\theta} < \theta_{RC} < \frac{\pi}{4}$. These prove the lemma. $\square$

We now express the two payoffs as functions of the measure θ_p . Headquarters' payoff is the first dimension of effort in (1),

$$\pi(\theta_p) = b(\theta_p) \cos \frac{\pi}{4} + \beta(\theta_p) \cos \theta_p,$$

and the manager's payoff is $v(\theta_p) = v(b(\theta_p), \beta(\theta_p))$, where $b(\theta_p)$ and $\beta(\theta_p)$ are the optimal bonuses from Lemma 1, evaluated at $\theta = |\theta_p - \frac{\pi}{4}|$.

Proof of Lemma 2. We characterize the two payoffs in the three regions from Lemma 1's proof.

Pure relational region. If $\theta \geq \theta_{RC}$, then $b = 1$ and $\beta = 0$. Hence $\pi(\theta_p) = \cos \frac{\pi}{4}$ and $v(\theta_p) = \frac{1}{2}$. Both payoffs are flat in this region.

Pure formal region. If $\theta \leq \hat{\theta}$, then $b = 0$ and $\beta = \cos \theta$. Hence

$$\pi(\theta_p) = \cos(\theta_p - \frac{\pi}{4}) \cos \theta_p \quad \text{and} \quad v(\theta_p) = \frac{1}{2} \cos^2(\theta_p - \frac{\pi}{4}).$$

Differentiating headquarters' payoff gives $\frac{d\pi}{d\theta_p} = -\sin(2\theta_p - \frac{\pi}{4})$. Under $r \leq \sin^2(\frac{\pi}{8})$, we have $\hat{\theta} \leq \frac{\pi}{8}$. Thus, in terms of θ_p , the pure formal region satisfies $\theta_p \in [\frac{\pi}{4} - \hat{\theta}, \frac{\pi}{4} + \hat{\theta}] \subseteq [\frac{\pi}{8}, \frac{3\pi}{8}]$. On this θ_p -interval, $\pi(\theta_p)$ decreases, and it strictly decreases on the interior $(\frac{\pi}{4} - \hat{\theta}, \frac{\pi}{4} + \hat{\theta})$. The manager's payoff is maximized at $\theta_p = \frac{\pi}{4}$ and strictly decreases as θ_p moves away from $\frac{\pi}{4}$.

Interior region. If $\hat{\theta} < \theta < \theta_{RC}$, then $b = 2 \left(1 - \frac{r}{\sin^2 \theta} \right)$ and $\beta = (1 - b) \cos \theta$. Substituting these bonuses into headquarters' payoff gives

$$\pi(\theta_p) = 2 \left(1 - \frac{r}{\sin^2(\theta_p - \frac{\pi}{4})} \right) \cos \frac{\pi}{4} + \left[1 - 2 \left(1 - \frac{r}{\sin^2(\theta_p - \frac{\pi}{4})} \right) \right] \cos(\theta_p - \frac{\pi}{4}) \cos \theta_p.$$

Differentiating with respect to θ_p gives

$$\frac{\partial \pi}{\partial \theta_p} = \frac{\sqrt{2}r}{\sin^2(\theta_p - \frac{\pi}{4})} + \sin(2\theta_p - \frac{\pi}{4}).$$

On the right interior interval $\theta_p \in (\frac{\pi}{4} + \hat{\theta}, \frac{\pi}{4} + \theta_{RC})$, both terms are strictly positive, so $\pi(\theta_p)$ strictly increases.

On the left interior interval, $\theta_p \in (\frac{\pi}{4} - \theta_{RC}, \frac{\pi}{4} - \hat{\theta})$, the derivative is positive at the left endpoint:

$$\left. \frac{\partial \pi}{\partial \theta_p} \right|_{\theta_p = \frac{\pi}{4} - \theta_{RC}} = \frac{\sqrt{2}}{2} + \sin(\frac{\pi}{4} - 2\theta_{RC}) > 0,$$

where the inequality uses $\theta_{RC} < \frac{\pi}{4}$. Moreover,

$$\frac{\partial^2 \pi}{\partial \theta_p^2} = 2 \cos(2\theta_p - \frac{\pi}{4}) + 2\sqrt{2}r \frac{\cos(\theta_p - \frac{\pi}{4})}{-\sin^3(\theta_p - \frac{\pi}{4})} > 0,$$

because both terms are strictly positive when $\theta_p \in (\frac{\pi}{4} - \theta_{RC}, \frac{\pi}{4} - \hat{\theta}) \subseteq (0, \frac{\pi}{4})$. Hence $\pi(\theta_p)$ strictly increases on $(\frac{\pi}{4} - \theta_{RC}, \frac{\pi}{4} - \hat{\theta})$.

For the manager, the non-reneging constraint binds in the interior region, so

$$v(\theta_p) = v^{\text{spot}} + rb = \frac{1}{2} \cos^2 \theta + 2r - \frac{2r^2}{\sin^2 \theta}.$$

Differentiating with respect to θ gives

$$\frac{dv}{d\theta} = \frac{\cos \theta}{\sin^3 \theta} (2r - \sin^2 \theta)(2r + \sin^2 \theta).$$

Because $\theta < \theta_{RC}$ implies $\sin^2 \theta < 2r$, this derivative is strictly positive. Thus the manager's payoff increases with the distortion θ : it strictly decreases in θ_p on $(\frac{\pi}{4} - \theta_{RC}, \frac{\pi}{4} - \hat{\theta})$ and strictly increases in θ_p on $(\frac{\pi}{4} + \hat{\theta}, \frac{\pi}{4} + \theta_{RC})$.

Combining the three regions gives Lemma 2. $\square$

Proof of Proposition 1. Headquarters chooses $\theta_p \in [\bar{\theta}, \frac{\pi}{2}]$ or withholds the measure. If headquarters withholds the measure, the manager relies on the relational contract, so headquarters obtains $\cos \frac{\pi}{4}$. Using Lemma 2 and its proof, we consider three cases.

Low capability ($\bar{\theta} \in (\frac{\pi}{4}, \frac{\pi}{2}]$). Every feasible measure yields at most $\cos \frac{\pi}{4}$, so withholding is optimal.

Moderate capability ($\bar{\theta} \in [\frac{\pi}{4} - \hat{\theta}, \frac{\pi}{4}]$). Headquarters' payoff rises as θ_p falls on $[\bar{\theta}, \frac{\pi}{4}]$. Because $\pi(\frac{\pi}{4}) = \cos \frac{\pi}{4}$, while measures above $\frac{\pi}{4}$ yield at most $\cos \frac{\pi}{4}$, $\theta_p^* = \bar{\theta}$.

High capability ($\bar{\theta} \in [0, \frac{\pi}{4} - \hat{\theta})$). Headquarters' payoff is maximized at $\frac{\pi}{4} - \hat{\theta}$, which is feasible. Hence $\theta_p^* = \frac{\pi}{4} - \hat{\theta}$. $\square$

Proof of Proposition 2. The result follows by substituting the optimal design from Proposition 1 into headquarters' payoff. $\square$

References

- Acemoglu, Daron, and Pascual Restrepo.** 2019. "Artificial Intelligence, Automation, and Work." In *The Economics of Artificial Intelligence: An Agenda.*, ed. Ajay Agrawal, Joshua Gans and Avi Goldfarb, 197–236. University of Chicago Press.
- Aghion, Philippe, and Jean Tirole.** 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105(1): 1–29.
- Athey, Susan C., Kevin A. Bryan, and Joshua S. Gans.** 2020. "The Allocation of Decision Authority to Human and Artificial Intelligence." *AEA Papers and Proceedings*, 110: 80–84.
- Baker, George, Robert Gibbons, and Kevin J Murphy.** 1994. "Subjective performance measures in optimal incentive contracts." *The Quarterly Journal of Economics*, 109(4): 1125–1156.
- Baker, George, Robert Gibbons, and Kevin J Murphy.** 1999. "Informal authority in organizations." *The Journal of Law, Economics, and Organization*, 15(1): 56–73.
- Baker, George, Robert Gibbons, and Kevin J Murphy.** 2002. "Relational Contracts and the Theory of the Firm." *The Quarterly Journal of Economics*, 117(1): 39–84.
- Baker, George, Robert Gibbons, and Kevin J Murphy.** 2023. "From Incentives to Control to Adaptation: Exploring Interactions between Formal and Relational Governance." *Journal of Institutional and Theoretical Economics*, 179(3–4): 500–529.
- Bolton, Patrick, and Mathias Dewatripont.** 2013. "Authority in Organizations: A Survey." In *The Handbook of Organizational Economics.*, ed. Robert Gibbons and John Roberts, 342–372. Princeton, NJ: Princeton University Press.
- Brynjolfsson, Erik, and Paul Milgrom.** 2013. "Complementarity in Organizations." In *The Handbook of Organizational Economics.*, ed. Robert Gibbons and John Roberts, 11–55. Princeton University Press.
- Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond.** 2025. "Generative AI at Work." *Quarterly Journal of Economics*, 140(2): 889–942.
- Brynjolfsson, Erik, Daniel Rock, and Chad Syverson.** 2021. "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies." *American Economic Journal: Macroeconomics*, 13(1): 333–372.

- Chandler, Alfred D.** 1977. *The Visible Hand: The Managerial Revolution in American Business*. Harvard University Press.
- Chen, Yvonne Jie, Jie Gong, Jin Li, and Zibo Zhao.** 2025. “Better Technology, Worse Motivation: GenAI’s Mediocrity Trap.” HKU Centre for AI, Management and Organization CAMO Working Paper 2025-01.
- Choe, Chongwoo, and Shingo Ishiguro.** 2022. “Relational Contracts and Hierarchy.” Monash University, Department of Economics Monash Economics Working Paper 2022-08.
- Dalton, Melville.** 1959. *Men Who Manage: Fusions of Feeling and Theory in Administration*. John Wiley & Sons.
- de la Garza, Alejandro.** 2019. “This AI Software Is ‘Coaching’ Customer Service Workers. Soon It Could Be Bossing You Around, Too.” *TIME*, <https://time.com/5610094/cogito-ai-artificial-intelligence/>.
- Fahn, Matthias, Jin Li, and Chang Sun.** 2026. “Toward a Bad Job Economy: AI Adoption, Agency Costs, and Job Design.” IZA@LISER Network IZA Discussion Paper 18574.
- Faure-Grimaud, Antoine, Jean-Jacques Laffont, and David Martimort.** 2003. “Collusion, delegation and supervision with soft information.” *The Review of Economic Studies*, 70(2): 253–279.
- Fei, Xiaotong.** 1992. *From the Soil: The Foundations of Chinese Society*. University of California Press. Translated by Gary G. Hamilton and Wang Zheng.
- Field, Emily, Bryan Hancock, and Bill Schaninger.** 2023. “Middle Managers Are the Heart of Your Company.” *McKinsey Quarterly*.
- Fong, Yuk-fai, and Jin Li.** 2017. “Information Revelation in Relational Contracts.” *Review of Economic Studies*, 84(1): 277–299.
- Holmström, Bengt, and Paul Milgrom.** 1991. “Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design.” *The Journal of Law, Economics, and Organization*, 7(Special Issue): 24–52.
- Ide, Enrique, and Eduard Talamàs.** 2025. “Artificial Intelligence in the Knowledge Economy.” *Journal of Political Economy*, 133(12): 3762–3800.

- Kellogg, Katherine C., Melissa A. Valentine, and Angèle Christin.** 2020. "Algorithms at Work: The New Contested Terrain of Control." *Academy of Management Annals*, 14(1): 366–410.
- Kerr, Steven.** 1975. "On the folly of rewarding A, while hoping for B." *Academy of Management Journal*, 18(4): 769–783.
- Kofman, Fred, and Jacques Lawarrée.** 1993. "Collusion in Hierarchical Agency." *Econometrica*, 61(3): 629–656.
- Leonardi, Paul, and Noshir Contractor.** 2018. "Better People Analytics: Measure Who They Know, Not Just Who They Are." *Harvard Business Review*, 96(6): 70–81.
- Li, Hanzhe, Jin Li, Arijit Mukherjee, and Luís Vasconcelos.** 2025. "Managing Loyalty in Hierarchical Organizations." Working Paper.
- Li, Jin, Feng Zhu, and Pascal Hua.** 2025. "Overcoming the Organizational Barriers to AI Adoption." *Harvard Business Review*.
- Li, Jin, Lingfang Ivy Li, Yiyang Zhou, and Yao Huang.** 2026. "Performance Evaluation and Employee Incentives in the AI Era: Dingdong's Transformation of VOC Assessment." Nano Case, HKU Centre for AI, Management and Organization. <https://camo.hku.hk/nanocase-dingdong/>.
- Li, Lingfang, Yangbo Chen, and Yi Liu.** 2025. "Dingdong (Cayman): Leader in Community Fresh-Food E-Commerce Business." Fudan University School of Management Case FDC-21003-1X-CE (updated version). <https://doi.org/10.12156/FUDAN.CASE202100302>.
- Mookherjee, Dilip.** 2013. "Incentives in Hierarchies." In *The Handbook of Organizational Economics*, ed. Robert Gibbons and John Roberts, 764–798. Princeton, NJ: Princeton University Press.
- Tirole, Jean.** 1986. "Hierarchies and bureaucracies: On the role of collusion in organizations." *The Journal of Law, Economics, and Organization*, 2(2): 181–214.
- Tong, Siliang, Nan Jia, Xueming Luo, and Zheng Fang.** 2021. "The Janus Face of Artificial Intelligence Feedback: Deployment versus Disclosure Effects on Employee Performance." *Strategic Management Journal*, 42(9): 1600–1631.
- Troya-Martinez, Marta, and Liam Wren-Lewis.** 2023. "Managing relational contracts." *Journal of the European Economic Association*, 21(3): 941–986.

Tsipursky, Gleb. 2026. "Middle Managers Will Make or Break AI Adoption." *Harvard Business Review*.

Supplemental Appendix to “Governing Informal Incentives: AI Capability and Performance Measure Design in Hierarchical Organizations”

YAO HUANG JIN LI LINGFANG (IVY) LI

Notation. We follow the main text. For a measure θ_p with distortion $\theta = |\theta_p - \frac{\pi}{4}|$, $\pi(\theta_p)$ and $v(\theta_p)$ denote headquarters' and the manager's payoffs under the manager's optimal contracts $b(\theta)$ and $\beta(\theta)$ (Lemma 1 and the Appendix of the main text). The thresholds are $\hat{\theta} = \arccos \sqrt{1-r}$ and $\theta_{RC} = \arccos \sqrt{1-2r}$, with $0 < r \leq \sin^2(\frac{\pi}{8})$. AI capability is $\bar{\theta}$, the most aligned feasible measure.

S1 Headquarters Maximizing Total Surplus

Suppose that, at the outset, headquarters chooses the performance measure and offers the manager a fixed transfer t , expressed as a per-period equivalent. The manager then decides whether to accept the offer or take an outside option worth $\bar{v} \geq 0$, also expressed as a per-period equivalent. If she accepts, the repeated game between the manager and the worker proceeds exactly as in the baseline model.

The manager accepts headquarters' offer if $t + v(\theta_p) \geq \bar{v}$. Headquarters sets the transfer so that this participation constraint binds, yielding $t = \bar{v} - v(\theta_p)$. The worker's participation constraint also binds. Total surplus generated within the hierarchy is

$$S(\theta_p) := \mathbb{E}[y_h + y_m \mid e] - c(e) = \pi(\theta_p) + v(\theta_p),$$

and headquarters' payoff is $\pi(\theta_p) - t = S(\theta_p) - \bar{v}$. Because $\bar{v}$ is constant, headquarters chooses θ_p to maximize $S(\theta_p)$.

Let $\tilde{\theta} := \frac{\pi}{4} - \frac{1}{2} \arctan\left(1 + \frac{\sqrt{2}}{2}\right) \approx 0.265$ and $\theta_S := \min\{\hat{\theta}, \tilde{\theta}\}$.

Lemma S1. *As the performance measure θ_p becomes more aligned with headquarters' goal, moving from $\frac{\pi}{2}$ toward 0, total surplus $S(\theta_p)$ changes as follows:*

- (i) Total surplus decreases as θ_p decreases from $\frac{\pi}{2}$ to $\frac{\pi}{4} + \hat{\theta}$.
- (ii) Total surplus strictly increases as θ_p decreases from $\frac{\pi}{4} + \hat{\theta}$ to $\frac{\pi}{4} - \theta_S$.
- (iii) Total surplus decreases as θ_p decreases from $\frac{\pi}{4} - \theta_S$ to 0.

Lemma S1 shows that total surplus is maximized at $\frac{\pi}{4} - \theta_S$, a measure below the manager's goal. Two forces limit how far headquarters aligns the measure. The first is the relational-contracting constraint from the baseline: below $\frac{\pi}{4} - \hat{\theta}$, the manager sustains a relational contract that pulls effort back toward her goal. The second comes from the total-surplus objective itself: even without a relational contract, moving the measure further toward headquarters' goal raises headquarters' payoff but lowers the manager's. The angle $\tilde{\theta}$ balances this second tradeoff, so headquarters holds back at $\frac{\pi}{4} - \min\{\hat{\theta}, \tilde{\theta}\}$.

Proposition S1. *Under the total-surplus objective, the optimal performance measure depends on the firm's AI capability:*

- (i) *For firms with low capability ($\bar{\theta} \in (\frac{\pi}{4}, \frac{\pi}{2}]$), headquarters implements no formal measure.*
- (ii) *For firms with moderate capability ($\bar{\theta} \in [\frac{\pi}{4} - \theta_S, \frac{\pi}{4}]$), headquarters implements $\theta_p^* = \bar{\theta}$.*
- (iii) *For firms with high capability ($\bar{\theta} \in [0, \frac{\pi}{4} - \theta_S)$), headquarters implements $\theta_p^* = \frac{\pi}{4} - \theta_S$.*

Total surplus under the optimal measure is constant as $\bar{\theta}$ decreases from $\frac{\pi}{2}$ to $\frac{\pi}{4}$, strictly increases as $\bar{\theta}$ decreases from $\frac{\pi}{4}$ to $\frac{\pi}{4} - \theta_S$, and is constant as $\bar{\theta}$ decreases from $\frac{\pi}{4} - \theta_S$ to 0.

Proposition S1 has the same three regimes as Proposition 1 in the main text. As AI capability improves, total surplus stays flat at low capability, rises at moderate capability, and plateaus at high capability. Only the held-back point moves. Because $\theta_S \leq \hat{\theta}$ as headquarters now internalizes the manager's payoff, headquarters holds back weakly more than in the baseline.

S2 Organizational Form

In this section headquarters can choose whether to retain or remove the managerial layer. If it retains the layer, it offers the manager the fixed transfer of Section S1. If it removes the layer, it contracts directly with the worker through formal contracting.

Proposition S2. *The optimal organizational form depends on AI capability $\bar{\theta}$ and the manager's outside option $\bar{v}$ as follows:*

- (i) *For sufficiently low values of $\bar{v}$, headquarters always retains the managerial layer.*
- (ii) *For intermediate values of $\bar{v}$, there is a cutoff $\bar{\theta}_D$ in the high-capability region. Headquarters retains the managerial layer when $\bar{\theta} > \bar{\theta}_D$ and removes the layer when $\bar{\theta} < \bar{\theta}_D$.*

(iii) For still higher values of $\bar{v}$, headquarters may remove the managerial layer even at lower levels of AI capability.

Proposition S2 shows when AI capability leads the firm to remove the managerial layer. If the manager's outside option is low, retaining her remains optimal even as AI capability becomes arbitrarily high. If her outside option is higher, headquarters initially retains the layer but removes it once AI capability becomes sufficiently high, as Figure S1 illustrates. To see why, notice that as AI capability improves, direct contracting becomes more valuable, while the payoff from retaining the managerial layer eventually plateaus. A higher outside option makes the manager more costly to retain, so sufficiently high AI capability leads headquarters to remove the layer.

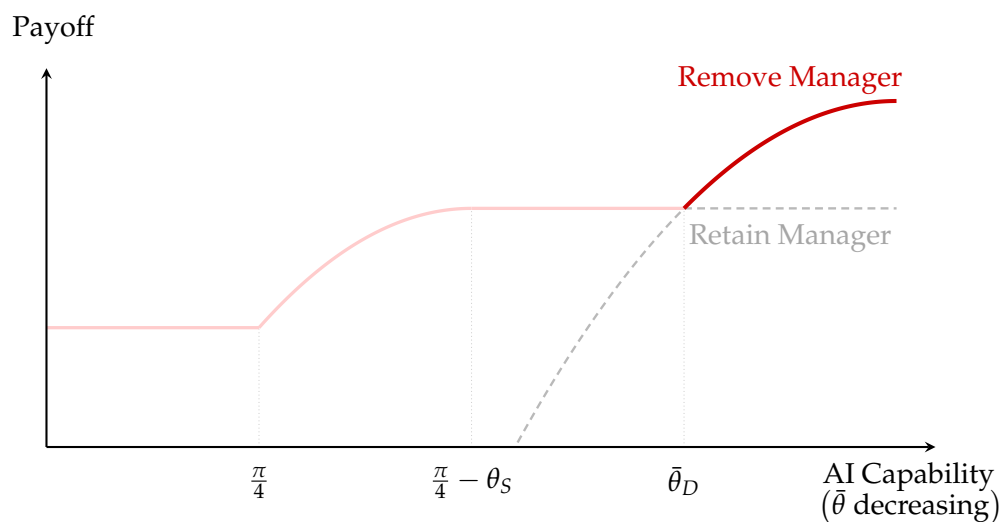


Figure S1: Organizational Form and the Return to AI Capability

Proof of Lemma S1. We use the bonuses and payoffs from the proofs of Lemmas 1 and 2 in the main text, and write $\theta = |\theta_p - \frac{\pi}{4}|$.

Pure relational region. If $\theta \geq \theta_{RC}$, then $b = 1$ and $\beta = 0$, so effort is $(\cos \frac{\pi}{4}, \sin \frac{\pi}{4})$ for every θ_p and total surplus is constant at $\cos \frac{\pi}{4} + \frac{1}{2}$.

Right interior region. If $\theta_p \in (\frac{\pi}{4} + \hat{\theta}, \frac{\pi}{4} + \theta_{RC})$, both $\pi(\theta_p)$ and $v(\theta_p)$ strictly increase in θ_p , hence so does S .

Pure formal region. If $\theta \leq \hat{\theta}$, then $b = 0$ and $\beta = \cos \theta$, so $S(\theta_p) = \cos \theta \cos \theta_p + \frac{1}{2} \cos^2 \theta$ and

$$\frac{dS}{d\theta_p} = -\sin(2\theta_p - \frac{\pi}{4}) - \frac{1}{2} \sin(2\theta_p - \frac{\pi}{2}).$$

On $[\frac{\pi}{4}, \frac{\pi}{4} + \hat{\theta}]$, $dS/d\theta_p$ is strictly negative, so S strictly increases as θ_p decreases to $\frac{\pi}{4}$. On $[\frac{\pi}{4} - \hat{\theta}, \frac{\pi}{4}]$,

$$\frac{d^2S}{d\theta_p^2} = -2 \cos(2\theta_p - \frac{\pi}{4}) - \sin(2\theta_p) < 0,$$

because $\hat{\theta} \leq \frac{\pi}{8}$ implies $\theta_p \in [\frac{\pi}{8}, \frac{\pi}{4}]$, where $\cos(2\theta_p - \frac{\pi}{4})$ and $\sin(2\theta_p)$ are both strictly positive. The first-order condition is $\tan(2\theta_p) = 1 + \frac{\sqrt{2}}{2}$, with unique solution $\theta_p = \frac{\pi}{4} - \tilde{\theta}$. By strict concavity, S is maximized on $[\frac{\pi}{4} - \hat{\theta}, \frac{\pi}{4}]$ at the point of that interval nearest $\frac{\pi}{4} - \tilde{\theta}$, which is $\frac{\pi}{4} - \theta_S$ because $\tilde{\theta} > 0$. S strictly increases as θ_p decreases to that point and decreases as θ_p decreases further.

Left interior region. If $\theta_p \in (\frac{\pi}{4} - \theta_{RC}, \frac{\pi}{4} - \hat{\theta})$, then $\theta = \frac{\pi}{4} - \theta_p$, and the proof of Lemma 2 in the main text gives

$$\frac{d\pi}{d\theta_p} = \frac{\sqrt{2}r}{\sin^2 \theta} + \sin(\frac{\pi}{4} - 2\theta) \quad \text{and} \quad \frac{dv}{d\theta_p} = \sin \theta \cos \theta - \frac{4r^2 \cos \theta}{\sin^3 \theta}.$$

These yield

$$\frac{dS}{d\theta_p} = \frac{\sqrt{2}r}{\sin^2 \theta} \left(1 - 2\sqrt{2}r \frac{\cos \theta}{\sin \theta} \right) + (\sin(\frac{\pi}{4} - 2\theta) + \frac{1}{2} \sin(2\theta)).$$

Both terms are positive. For the first, because $r = \sin^2 \hat{\theta}$ and $\theta > \hat{\theta}$ on this region,

$$2\sqrt{2}r \frac{\cos \theta}{\sin \theta} < 2\sqrt{2} \sin \hat{\theta} \cos \hat{\theta} = \sqrt{2} \sin(2\hat{\theta}) \leq 1,$$

where the last inequality uses $\hat{\theta} \leq \frac{\pi}{8}$. For the second, expanding $\sin(\frac{\pi}{4} - 2\theta)$ gives $\frac{\sqrt{2}}{2} \cos(2\theta) - \frac{\sqrt{2}-1}{2} \sin(2\theta)$, which for $2\theta < \frac{\pi}{2}$ is positive if and only if $\tan(2\theta) < 2 + \sqrt{2}$. This holds because $\sin^2 \theta_{RC} = 2r \leq 1 - \frac{\sqrt{2}}{2}$ gives $\theta < \theta_{RC} \leq \arcsin \sqrt{1 - \frac{\sqrt{2}}{2}} \approx 0.572$, while $\frac{1}{2} \arctan(2 + \sqrt{2}) \approx 0.643$. So S strictly increases in θ_p on this region. The four regions and the continuity of S give the lemma.

□

Proof of Proposition S1. Withholding the measure leaves the manager with the relational contract alone, so total surplus is $\cos \frac{\pi}{4} + \frac{1}{2}$. This equals $S(\frac{\pi}{4})$, because $\theta_p = \frac{\pi}{4}$ induces the same effort, and also equals $S(\frac{\pi}{2})$. By Lemma S1, every measure weakly above $\frac{\pi}{4}$ therefore yields at most $S(\frac{\pi}{4})$, and total surplus strictly increases as θ_p decreases to $\frac{\pi}{4} - \theta_s$ and decreases thereafter. Hence, when $\bar{\theta} > \frac{\pi}{4}$, withholding is optimal. When $\bar{\theta} \in [\frac{\pi}{4} - \theta_s, \frac{\pi}{4}]$, surplus strictly increases as θ_p decreases to $\bar{\theta}$, so $\theta_p^* = \bar{\theta}$. When $\bar{\theta} < \frac{\pi}{4} - \theta_s$, the measure $\frac{\pi}{4} - \theta_s$ is feasible and strictly better than any other measure, so $\theta_p^* = \frac{\pi}{4} - \theta_s$. The final statement follows by substituting θ_p^* into S and applying Lemma S1(ii). $\square$

Proof of Proposition S2. Let $S^* := S(\frac{\pi}{4} - \theta_s)$, and let $S^H(\bar{\theta})$ denote the highest total surplus generated within the hierarchy. Proposition S1 implies

$$S^H(\bar{\theta}) = \begin{cases} \cos \frac{\pi}{4} + \frac{1}{2}, & \bar{\theta} \in (\frac{\pi}{4}, \frac{\pi}{2}], \\ S(\bar{\theta}), & \bar{\theta} \in [\frac{\pi}{4} - \theta_s, \frac{\pi}{4}], \\ S^*, & \bar{\theta} \in [0, \frac{\pi}{4} - \theta_s]. \end{cases}$$

Headquarters' payoff from retaining the managerial layer is $S^H(\bar{\theta}) - \bar{v}$.

After removing the layer, headquarters offers the worker a fixed salary and a formal bonus β based on θ_p . The worker's participation constraint binds, so headquarters chooses β to maximize $\beta \cos \theta_p - \frac{1}{2}\beta^2$. It therefore sets $\beta = \cos \theta_p$ and chooses the most aligned feasible measure $\theta_p = \bar{\theta}$. Direct contracting consequently yields $S^D(\bar{\theta}) = \frac{1}{2} \cos^2 \bar{\theta}$.

Headquarters retains the managerial layer whenever $S^H(\bar{\theta}) - S^D(\bar{\theta}) \geq \bar{v}$. In the low-capability region, this surplus advantage of the hierarchy exceeds $\frac{1}{\sqrt{2}} + \frac{1}{4}$. In the moderate-capability region,

$$S^H(\bar{\theta}) - S^D(\bar{\theta}) = \frac{1}{2\sqrt{2}} + \frac{\sqrt{2}+1}{4} \sin(2\bar{\theta}) + \frac{\sqrt{2}-1}{4} \cos(2\bar{\theta}),$$

which is concave in $\bar{\theta}$ on this interval, so its minimum occurs at an endpoint. The smallest surplus advantage of the hierarchy before reaching the plateau is therefore

$$\Delta := \min \left\{ \frac{1}{\sqrt{2}} + \frac{1}{4}, S^* - \frac{1}{2} \cos^2 \left(\frac{\pi}{4} - \theta_s \right) \right\}.$$

Both terms defining Δ are strictly greater than $S^* - \frac{1}{2}$. For the second term, this follows from $\theta_s < \frac{\pi}{4}$. For the first, the optimal measure induces a pure formal contract, so

$$S^* = \cos \theta_s \cos \left(\frac{\pi}{4} - \theta_s \right) + \frac{1}{2} \cos^2 \theta_s \leq 1 + \frac{1}{2\sqrt{2}} < \frac{1}{\sqrt{2}} + \frac{3}{4}.$$

The first inequality is because of $\cos \theta_S \cos(\frac{\pi}{4} - \theta_S) \leq \frac{1}{2}(1 + \frac{1}{\sqrt{2}})$. Hence $(S^* - \frac{1}{2}, \Delta)$ is nonempty.

If $\bar{v} \leq S^* - \frac{1}{2}$, the surplus advantage of the hierarchy is at least $\bar{v}$ at every capability level. Headquarters therefore always retains the managerial layer, proving part (i).

Now suppose $S^* - \frac{1}{2} < \bar{v} < \Delta$. Headquarters retains the layer before the plateau. On the plateau, its surplus advantage decreases from $S^* - \frac{1}{2} \cos^2(\frac{\pi}{4} - \theta_S)$ to $S^* - \frac{1}{2}$ as $\bar{\theta}$ falls to zero. It therefore removes the layer once $\bar{\theta}$ falls below $\bar{\theta}_D(\bar{v}) = \arccos \sqrt{2(S^* - \bar{v})}$, where $\bar{\theta}_D(\bar{v}) \in (0, \frac{\pi}{4} - \theta_S)$. This proves part (ii).

Finally, if $\bar{v} > \Delta$, direct contracting becomes optimal at some capability levels before the hierarchy reaches the plateau. As $\bar{v}$ increases, the set of capability levels at which headquarters removes the layer expands, and for $\bar{v} > \cos \frac{\pi}{4} + \frac{1}{2}$ headquarters removes the layer at every capability level. This proves part (iii). $\square$